



HAL
open science

Study of DNA methylation modifications: from dynamics during the dedifferentiation into breast cancer stem cells, to the development of the R-based tool ABSP, analysis of bisulfite sequencing PCR

Marie Denoulet

► **To cite this version:**

Marie Denoulet. Study of DNA methylation modifications: from dynamics during the dedifferentiation into breast cancer stem cells, to the development of the R-based tool ABSP, analysis of bisulfite sequencing PCR. Cancer. Université de Lille, 2022. English. NNT : 2022ULILS105 . tel-04762929

HAL Id: tel-04762929

<https://theses.hal.science/tel-04762929v1>

Submitted on 1 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LILLE
ÉCOLE DOCTORALE BIOLOGIE-SANTÉ DE LILLE
LABORATOIRE CANTHER UMR9020 CNRS – UMR1277 INSERM

THÈSE DE DOCTORAT

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE LILLE
Aspects moléculaires et cellulaires de la biologie

**STUDY OF DNA METHYLATION MODIFICATIONS:
FROM DYNAMICS DURING THE DEDIFFERENTIATION INTO
BREAST CANCER STEM CELLS, TO THE DEVELOPMENT OF THE
R-BASED TOOL ABSP, ANALYSIS OF BISULFITE SEQUENCING PCR**

Présentée et soutenue publiquement par

Marie Denoulet

Le 31 octobre 2022

COMPOSITION DU JURY :

PRÉSIDENTE DU JURY	Pr. Xuefen Le Bourhis	Laboratoire Hétérogénéité, Plasticité et Résistance aux Thérapies des Cancers (CANTHER), Université de Lille
RAPPORTEURS	Pr. Marc Cuggia	Laboratoire Traitement du Signal et de l'Image (LTSI), Université de Rennes
	Dr. Thierry Virolle	Institut de Biologie Valrose (iBV), Université de Nice
EXAMINATRICES	Dr. Michelina Plateroti	Laboratoire Interface de Recherche Fondamentale et Appliquée en Cancérologie (IRFAC), Université de Strasbourg
	Dr. Audrey Vincent	Laboratoire Hétérogénéité, Plasticité et Résistance aux Thérapies des Cancers (CANTHER), Université de Lille
DIRECTEUR DE THÈSE	Dr. Chann Lagadec	Laboratoire Hétérogénéité, Plasticité et Résistance aux Thérapies des Cancers (CANTHER), Université de Lille

Remerciements

Merci à l'**Université de Lille** qui a financé mes trois premières années de thèse, ainsi qu'à la **Ligue contre le cancer** qui m'a permis de poursuivre ce travail avec le financement d'une quatrième année.

Merci au Professeur **Marc Cuggia**, au Docteur **Thierry Virolle**, à la Docteure **Michelina Plateroti** et à la Docteure **Audrey Vincent** d'avoir accepté de lire ce manuscrit et d'évaluer mes travaux. Votre participation à mon jury de thèse m'honore, et je vous remercie pour le temps que vous y avez consacré. Et merci à la Docteure **Michelina Plateroti** et à la Docteure **Audrey Vincent** pour votre participation à mes comités de suivi individuels et pour tous les conseils prodigués.

Merci à la Professeure **Xuefen Le Bourhis**, directrice de l'équipe Plasticité Cellulaire et Cancer au sein du laboratoire CANTHER. Merci de m'avoir accueillie pour mon master 2 puis pour ma thèse. Votre intérêt pour mon travail et vos conseils m'ont beaucoup appris, particulièrement lors de nos réunions d'équipe.

Merci à **Chann** pour ton encadrement depuis mon master 2, merci de m'avoir fait confiance durant ces 5 années au laboratoire. Ta disponibilité, ton enthousiasme, et tes conseils — toujours les bons mots aux bons moments — sont de précieux atouts qui font de toi un admirable directeur de thèse.

Merci à la plateforme **Bilille**, à **Martin Figeac** et à **Franck Bonardi**, pour leur aide sur les analyses et au delà, leur avis et expertise ont été d'une aide précieuse pour moi.

Merci à **Nathalie** et **Émilie** pour tous ces moments passés auprès des cytomètres pendant de longues heures de tri, vous m'avez toujours accueillie avec bonne humeur et conseillée au mieux.

Merci à **Robert-Alain**, **Éric** et **Pierre-Olivier** pour vos retours, discussions et conseils, lors de nos réunions et en particulier lors de la préparation au concours, vous avez à coeur de voir vos étudiants réussir.

Merci à **Karine** qui m'a épaulée en master 2 et au début de ma thèse, merci pour ton aide, ton optimisme et ta bonne humeur, même quand les expériences ne donnaient rien.

Merci à **Ihsan**, de m'avoir formée, conseillée, et soutenue, ton sourire radieux et ta gentillesse ont été une véritable chance pour l'équipe.

Merci à **Valérie** et **Isabelle** pour leur présence et leur aide précieuse dans la vie quotidienne du laboratoire.

Merci à **Justine** et **Nadège**, pour votre implication, formation et soutien, merci de m'avoir accompagnée dans mes débuts de thèse et pour les bons moments passés au sein de la CSC team.

Merci aux autres doctorants passés avant moi, **Barbara**, **Romain**, **Jordan** et **Clément**, vous avez contribué à l'ambiance joviale et d'entraide entre nous, que ce soit au laboratoire et en dehors.

Merci aux futurs doctorants pour votre implication dans la vie du labo lors de l'année de master 2, réussie avec brio sans exception, merci particulièrement à **Flavie** et **Anaïs**, vous allez prendre la relève de la CSC team, je vous souhaite le meilleure pour vos thèses à venir.

Merci à tous les doctorants de l'équipe, **Ludivine**, **Julien**, **Alexandre**, **Amina**, **Évodie**, **Sarah**, **Marie**, **Mayar** et **Geoffrey**, pour tous les bons moments passés ensemble, les discussions de couloir, les sorties, les souvenirs à Deauville...

Merci à **Mathilde**, 4 ans partagés à tes côtés dans le bureau de la CSC team, toujours une oreille tendue et un sourire, tu es l'une des plus belle rencontre de cette thèse et je te souhaite le meilleure pour ta fin de thèse et la suite de tes aventures.

Merci à **ma famille** et à **ma belle-famille**, je suis entourée de personnes formidables qui ont toujours été présents pour moi.

Merci à **Géry**, pour ton soutien incroyable, ta patience, tes corrections, et merci de partager ma vie au quotidien.

ABSTRACT

STUDY OF DNA METHYLATION MODIFICATIONS: FROM DYNAMICS DURING THE DEDIFFERENTIATION INTO BREAST CANCER STEM CELLS, TO THE DEVELOPMENT OF THE R-BASED TOOL ABSP, ANALYSIS OF BISULFITE SEQUENCING PCR

Cancer stem cells (CSCs) form a tumoral subpopulation characterized by self-renewal abilities, pluripotency, therapeutic resistance mechanisms, and tumor initiation capacities, and are therefore a major cause of cancer recurrence after treatments. Moreover, the non-cancer stem cells (non-CSCs) are able to dedifferentiate into CSCs, in response to stress, especially to antitumor treatments such as radiotherapy, thus reinforcing the therapeutic resistance of cancer. In addition, epigenetic marks such as DNA methylation are known to contribute to the regulation of stemness properties and could be involved in the reacquisition of a CSC phenotype.

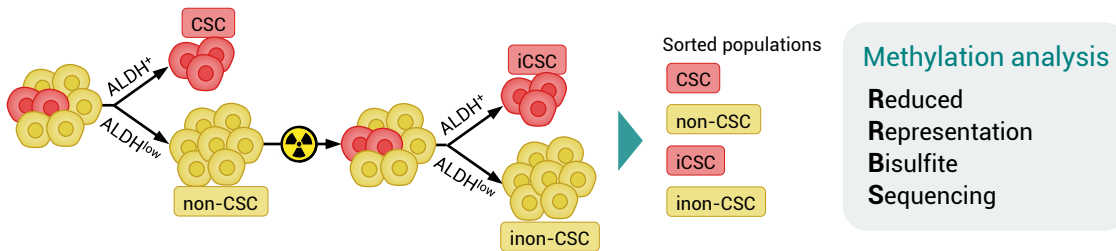
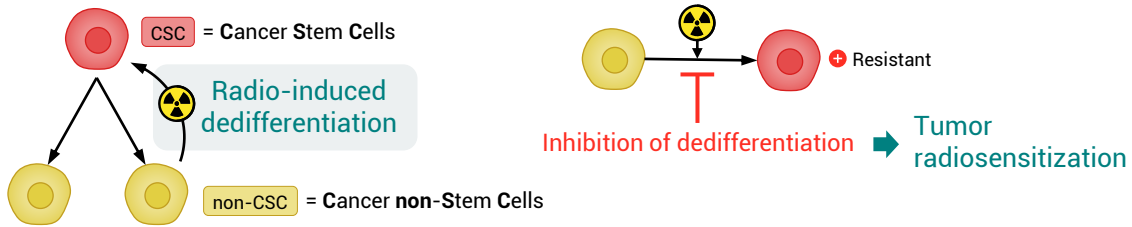
To evaluate DNA methylation modifications occurring throughout the radio-induced dedifferentiation of non-CSCs into CSCs in the breast cancer model, a Reduced Representation Bisulfite Sequencing (RRBS) analysis of the different tumor subpopulations was carried out. The analysis of RRBS data led to the identification of over 2,000 Differentially Methylated Regions (DMRs) undergoing methylation changes from non-CSC to radio-induced CSC. Among them, 35 present a methylation profile across the populations consistent with a potential contribution to radio-induced dedifferentiation. Five regions, associated with the FSCN1, CHRNA6, CDH7, CD9, and PRKAR1B genes, were selected for further validation. Genes regulated by these methylation changes could serve as new therapeutic targets to specifically inhibit the non-CSC to CSC phenotypic switch and prevent the enrichment in CSCs, reducing the risk of cancer relapse.

To validate identified methylation differences, the Bisulfite Sequencing PCR (BSP) method was chosen as it is the most convenient and accessible technique to quantify locus-specific methylation levels. Due to a lack of efficient tools to analyze BSP results from both approaches (direct-BSP and cloning-BSP), the ABSP R-based tool, standing for Analysis of Bisulfite Sequencing PCR, was developed. This tool provides a complete, automated, and user-friendly workflow to compute methylation percentages and compare methylation differences between samples. ABSP is available for download, along with associated data, at <https://github.com/ABSP-methylation-tool/ABSP>. Altogether, this work highlights the importance of DNA methylation within CSC plasticity and the room for tools to improve its analysis.

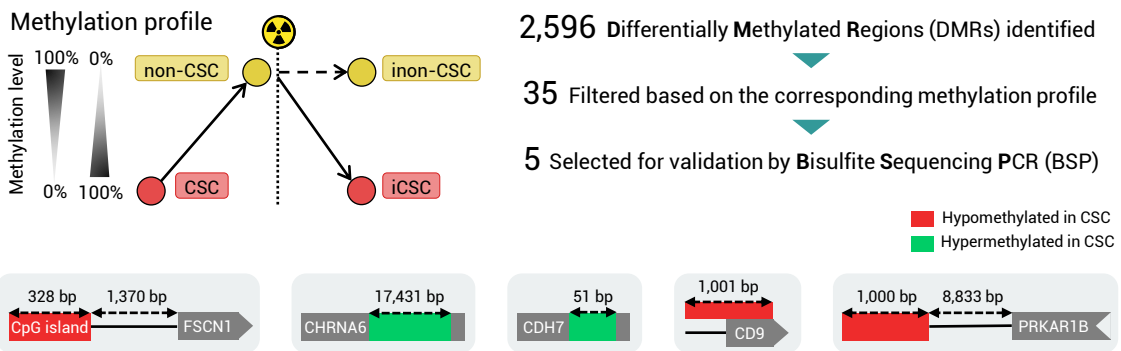
KEYWORDS: *breast cancer, radiotherapy, cancer stem cells, DNA methylation, bisulfite sequencing, R language*

STUDY OF DNA METHYLATION MODIFICATIONS:

from Dynamics during the Dedifferentiation into Breast Cancer Stem Cells



Identification of regions potentially involved in radio-induced dedifferentiation

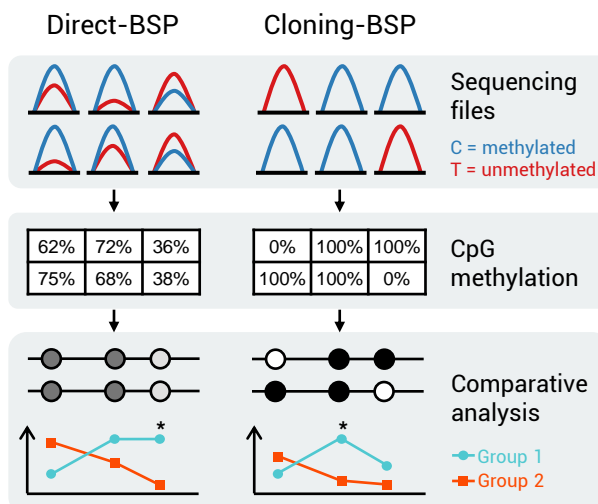


to the Development of the R-based Tool ABSP, Analysis of Bisulfite Sequencing PCR



Analysis of Bisulfite Sequencing PCR

- 1 Complete Workflow**
From raw data to comparative analysis
- 2 Fully Automated and user-friendly**
Built-in shiny app web interface
- 3 Analysis of both BSP approaches**
Direct-BSP & Cloning-BSP
- 4 Accessible and flexible**
R programming language



RÉSUMÉ

ÉTUDE DES MODIFICATIONS DE MÉTHYLATION DE L'ADN : DES DYNAMIQUES AU COURS DE LA DÉDIFFÉRENCIATION EN CELLULES SOUCHES CANCÉREUSES DE SEIN, AU DÉVELOPPEMENT DE L'OUTIL ABSP, *Analysis of Bisulfite Sequencing PCR*, SOUS R

Les Cellules Souches Cancéreuses (CSC) forment une sous-population tumorale caractérisée par des capacités d'auto-renouvellement, de pluripotence, d'initiation tumorale et présentent une résistance thérapeutique accrue. Elles sont donc une cause majeure de récurrence du cancer. De plus, les cellules cancéreuses non-souches sont capables de se différencier en CSC en réponse à un stress, notamment aux traitements anticancéreux comme la radiothérapie, renforçant ainsi la résistance thérapeutique de la tumeur. Nous avons fait l'hypothèse que les marques épigénétiques telles que la méthylation de l'ADN, connues comme contribuant à la régulation des propriétés souches, seraient impliquées dans la réacquisition d'un phénotype CSC.

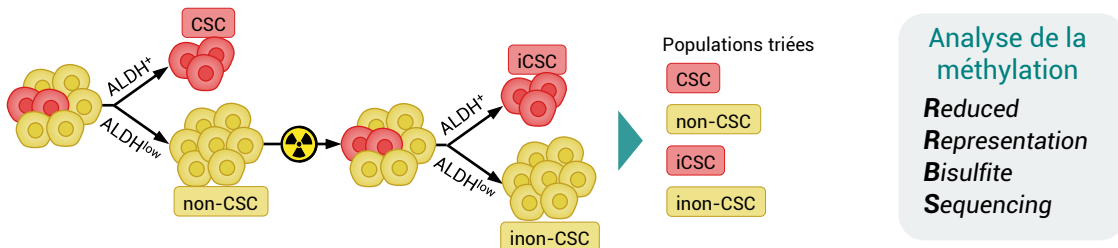
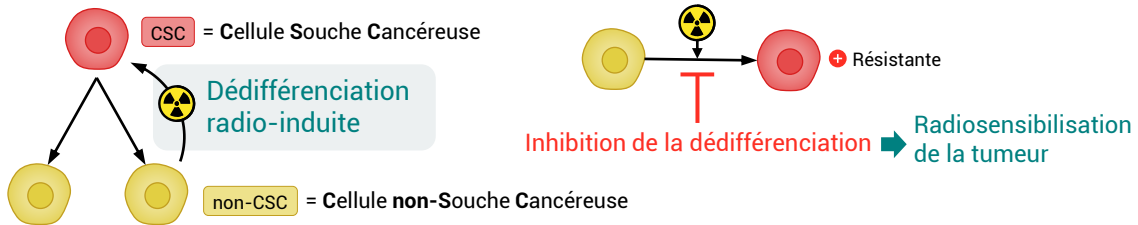
Afin d'évaluer les modifications de méthylation de l'ADN au cours de la différenciation radio-induite des cellules non-CSC en CSC dans le modèle de cancer du sein, une analyse de *Reduced Representation Bisulfite Sequencing* (RRBS) des différentes sous-populations tumorales a été réalisée. Cette analyse a permis d'identifier plus de 2 000 régions différentiellement méthylées (DMR) subissant des changements de méthylation entre les états non-CSC et CSC radio-induit. Nous avons retenu 35 DMR présentant un profil de méthylation cohérent avec une potentielle contribution à la différenciation radio-induite. Cinq d'entre elles, associées aux gènes *FSCN1*, *CHRNA6*, *CDH7*, *CD9* et *PRKAR1B*, ont été sélectionnées pour validation complémentaire. Les gènes régulés par ces changements de méthylation pourraient servir de nouvelles cibles thérapeutiques afin d'inhiber spécifiquement la conversion phénotypique de non-CSC à CSC et prévenir un enrichissement de la tumeur en CSC, réduisant ainsi le risque de rechute du cancer.

Pour valider les différences de méthylation observées en RRBS, la méthode de *Bisulfite Sequencing PCR* (BSP) a été choisie pour son accessibilité et son efficacité à quantifier les niveaux de méthylation d'un locus spécifique. En raison de l'absence d'outils à ce jour permettant d'analyser efficacement et de manière automatisée les résultats de BSP, provenant des deux approches de BSP (*direct-BSP* et *cloning-BSP*), nous avons donc fait le choix de développer sous R un nouvel outil, ABSP pour *Analysis of Bisulfite Sequencing PCR*. ABSP fournit une analyse complète, automatisée et accessible pour calculer les pourcentages de méthylation et comparer les différences de méthylation entre échantillons. ABSP et ses données associées sont téléchargeables à l'adresse <https://github.com/ABSP-methylation-tool/ABSP>. Ainsi, ce travail a mis en lumière l'importance de la méthylation de l'ADN dans la plasticité du phénotype souche cancéreux et le potentiel d'amélioration des outils d'analyse.

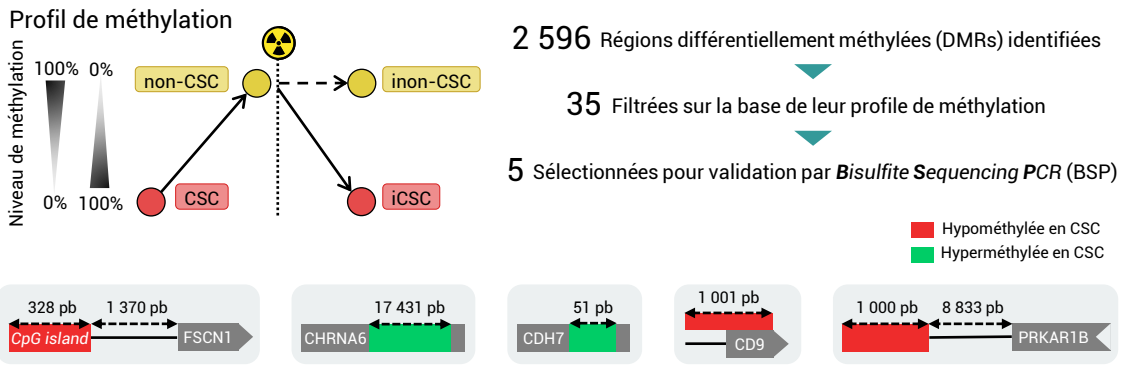
MOTS CLÉS : cancer du sein, radiothérapie, cellules souches cancéreuses, méthylation de l'ADN, séquençage bisulfite, langage R

ÉTUDE DES MODIFICATIONS DE MÉTHYLATION DE L'ADN:

des Dynamiques au cours de la Dédifférenciation en Cellules Souches Cancéreuses de Sein



Identification de régions potentiellement impliquées dans la dédifférenciation radio-induite

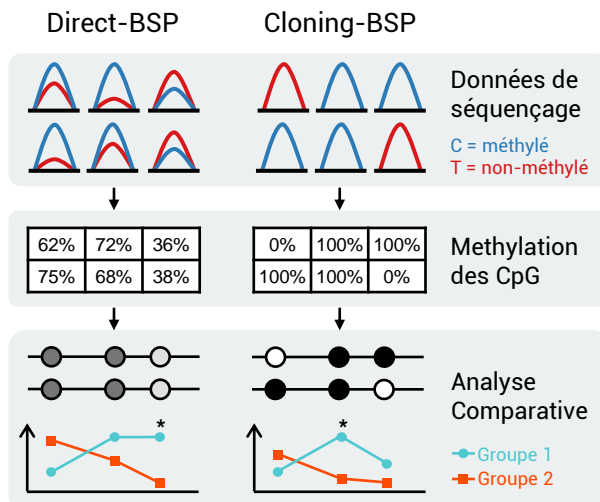


au Développement de l'Outil ABSP, Analysis of Bisulfite Sequencing PCR, sous R



Analysis of Bisulfite Sequencing PCR

- 1 Analyse complète**
Des données brutes à l'analyse comparative
- 2 Automatisé et facile à utiliser**
Interface web shiny intégrée
- 3 Analyse des deux approches**
Direct-BSP & Cloning-BSP
- 4 Accessible et flexible**
Langage de programmation R



SYNTHÈSE

ÉTUDE DES MODIFICATIONS DE MÉTHYLATION DE L'ADN : DES DYNAMIQUES AU COURS DE LA DÉDIFFÉRENCIATION EN CELLULES SOUCHES CANCÉREUSES DE SEIN, AU DÉVELOPPEMENT DE L'OUTIL ABSP, *Analysis of Bisulfite Sequencing PCR*, SOUS R

1 CONTEXTE

CELLULES SOUCHES CANCÉREUSES ET PLASTICITÉ PHÉNOTYPIQUE

Ces dernières années, le modèle stochastique d'organisation tumorale a été remplacé par un modèle alternatif basé sur l'organisation hiérarchique des cellules tumorales, faisant émerger le concept de cellules souches cancéreuses (CSC). Les CSC constituent une sous-population tumorale caractérisée par une résistance accrue aux thérapies, une capacité à s'autorenouveler et à régénérer une nouvelle tumeur, ce qui en fait donc une cause majeure des rechutes de cancers. De plus, il a été montré que des traitements anti-cancéreux comme la radiothérapie, induisait la dédifférenciation de cellules non-CSC en CSC dans le cancer du sein, ce qui conduit à un enrichissement en cellules résistantes au sein de la tumeur. Il est donc indispensable de développer de nouvelles approches permettant de prévenir la dédifférenciation afin de sensibiliser les tumeurs aux thérapies. Lors de cette conversion phénotypique, des changements transcriptomiques sont observés, tels que la réexpression des facteurs de pluripotence OCT4, SOX2 et NANOG. Ces modifications de programmes géniques pourraient donc être régulés par des mécanismes épigénétiques, comme la méthylation de l'ADN, pour permettre la réacquisition de propriétés de cellules souches.

MÉTHYLATION DE L'ADN ET BISULFITE SEQUENCING PCR (BSP)

La méthylation de l'ADN est le processus épigénétique le plus étudié chez les mammifères. Ce mécanisme se traduit par l'ajout d'un groupement méthyle sur des cytosines suivies d'une guanine, nommés sites CpG. La méthylation de l'ADN participe à la modulation l'architecture de la chromatine, notamment par la présence de régions denses en CpG, appelées *CpG islands*, ce qui permet la régulation de l'expression de gènes.

Diverses méthodes permettent de quantifier la méthylation de l'ADN. Parmi elles, le traitement de l'ADN au bisulfite convertit les cytosines (C) non méthylées en uraciles (U), alors que les cytosines méthylées (mC) restent cytosines (mC). Par amplification PCR, les uraciles (U) sont remplacés par des thymines (T), car tout deux complémentaires à l'adénine (A). Ainsi, après séquençage, la comparaison de la séquence convertie au bisulfite avec la séquence génomique originale permet de distinguer les statuts de méthylation de chaque site CpG. Pour un site CpG donné, la détection d'une

base C correspond donc au statut méthylé et d'une base T au statut non méthylé. Cette méthode permet donc de quantifier les pourcentages de méthylation d'une région spécifique au sein d'un mélange de molécules d'ADN. Cette technique est quantitative, accessible et offre un large spectre d'utilisation : de l'approche préliminaire à la validation de résultats obtenus à l'échelle du génome (RRBS/WGBS) en passant par l'analyse d'un grand nombre d'échantillons différents comme pour des études de cohortes. Deux types de BSP peuvent être distingués en fonction l'ADN séquencé, nommés *direct-BSP* et *cloning-BSP*. ① Dans la cas d'une approche *direct-BSP*, le séquençage de l'ensemble des amplicons permet une estimation semi-quantificative du pourcentage de méthylation d'un CpG dans la population totale des ADN. ② Dans la cas d'une approche *cloning-BSP*, le clonage des amplicons de PCR et le séquençage de clones individuels permet d'obtenir le statut de méthylation d'un CpG d'une molécule d'ADN (provenant d'un unique clone). Le pourcentage au sein de la population à chaque position est quantifié en calculant le ratio de clones méthylés et non-méthylés.

Actuellement, il n'existe qu'un seul outil permettant de calculer les niveaux de méthylation à partir de données de *direct-BSP*, appelé *ESME (Epigenetic Sequencing Methylation analysis)* et développé en 2004 par [Lewin et al. \(2004\)](#). Celui-ci nécessite un système d'exploitation Linux, ce qui n'est pas idéal pour une utilisation par des biologistes. Il est également limité uniquement au calcul de pourcentages de méthylation de données de *direct-BSP*, sans visualisation ou analyse statistique.

2 OBJECTIFS

IMPLICATION DE LA MÉTHYLATION DE L'ADN DANS LA DÉDIFFÉRENCIATION RADIO-INDUITE EN CSC

Dans une première partie, ce projet a pour but de déterminer l'implication des enzymes épigénétiques tels que les ADN méthyltransférases dans la dédifférenciation radio-induite de cellules non-CSC en CSC et d'identifier de potentiels nouveaux acteurs participant à ce processus. Nous avons émis l'hypothèse que des changements de méthylation de l'ADN intervenaient au cours de ce processus afin de modifier l'expression de certains gènes clés nécessaires à la réacquisition de propriétés de cellule souche. De ce fait, une analyse globale de changements de niveaux de méthylation de l'ADN permettrait d'identifier de nouveaux gènes et mécanismes impliqués dans la conversion phénotypique de non-CSC à CSC. A long terme, la validation de ces nouvelles cibles permettrait ainsi de pouvoir prévenir la dédifférenciation en CSC, pour réduire l'enrichissement de la tumeur en CSC plus résistantes après radiothérapie, afin de radiosensibiliser les tumeurs et diminuer le risque de rechute.

DÉVELOPPER UN OUTIL POUR ANALYSER LES DONNÉES DE BSP EFFICACEMENT ET DE MANIÈRE AUTOMATISÉE

Dans une deuxième partie, l'objectif est de développer un outil capable d'analyser à la fois des données de *direct-BSP* et de *cloning-BSP*, de manière automatisée et accessible aux chercheurs en biologie, et de manière complète, des données brutes jusqu'à l'analyse comparative, afin de déterminer les différences significatives de méthylation entre plusieurs groupes d'échantillons.

3 MÉTHODE

Pour connaître la contribution des enzymes épigénétiques dans la dédifférenciation radio-induite, l'expression des enzymes DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, et EZH2 a été mesurée au cours du temps après radiothérapie dans les cellules de cancer de sein SUM159PT. Ensuite, la contribution des ADN méthyltransférases DNMT1 et DNMT3B dans la dédifférenciation radio-induite a été évaluée à l'aide d'une inhibition par siRNA.

Afin d'identifier les changements de méthylation de l'ADN au niveau global, une analyse de *reduced representation bisulfite sequencing* (RRBS) a été menée sur différentes populations. En effet, les cellules SUM159PT ont été triées une première fois pour isoler les population CSC et non-CSC. La population de non-CSC a ensuite été irradiée pour induire leur dédifférenciation. Enfin, les cellules irradiées ont de nouveau été triées 5 jours après irradiation pour isoler les CSC induites (iCSC) et les non-CSC irradiées (inon-CSC). La comparaison des données de méthylation entre ces différentes populations a ensuite été réalisée par analyse bioinformatique sous R afin d'identifier des régions différenciellement méthylées (DMR).

L'outil ABSP a été développé sous le langage de programmation R, avec l'environnement de développement RStudio. Il intègre une application *shiny* qui s'ouvre sur un navigateur web et fait intervenir des scripts R *markdown* permettant de générer des rapports d'analyse sous format HTML. ABSP est divisé en deux analyses, l'une calcule les pourcentages de méthylation de chaque échantillon unique, l'autre regroupe tous les échantillons pour réaliser une analyse comparative et déterminer les différences de méthylation statistiquement significatives.

4 RÉSULTATS

4.1 DYNAMIQUES DE MÉTHYLATION DE L'ADN AU COURS DE LA DÉDIFFÉRENCIATION RADIO-INDUITE

IMPLICATION DES ENZYMES ÉPIGÉNÉTIQUES DANS LA DÉDIFFÉRENCIATION RADIO-INDUITE

L'analyse de l'expression des enzymes DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, et EZH2 après irradiation n'a pas permis de révéler des changements d'expression significatifs de ces enzymes, et donc leur implication dans la dédifférenciation n'a pas pu être démontré. En effet, une approche globale de mesure d'expression à certains temps déterminés après irradiation ne permettait pas de voir des changements d'expression transitoires liés aux événements de dédifférenciation rares (<5% de CSC induites 5 jours après irradiation), asynchrones et étalés dans le temps.

La transfection des cellules avec les siRNA a aboli l'augmentation du pourcentage de CSC après radiothérapie, à la fois dans la condition contrôle transfectée avec un siRNA contrôle et dans les conditions transfectées avec les siRNA ciblant DNMT1 et DNMT3B. Ces résultats indiquent que la transfection en elle-même est responsable de la diminution de la dédifférenciation. Ainsi, la participation de ces enzymes dans ce processus de conversion phénotypique n'a pas pu être déterminée.

IDENTIFICATION DE RÉGIONS DIFFÉRENTIELLEMENT MÉTHYLÉES

L'analyse des données de RRBS sur les populations CSC, non-CSC, iCSC, et inon-CSC a permis l'identification de 2596 régions différentiellement méthylées au cours de la dédifférenciation. Ces régions ont été filtrées pour ne garder que celles ayant un profil de méthylation correspondant à une régulation de la méthylation au cours de la dédifférenciation. 35 régions filtrées ont donc été identifiées comme potentiellement impliquées dans cette conversion d'un état non-CSC à CSC. Parmi ces 35 régions, 5 ont été retenues pour être validées à l'aide d'une méthode plus spécifique et précise, par exemple par BSP. ① Une région *CpG island* de 328 pb hypométhylée dans les CSC comparé aux non-CSC est localisée à 1370 pb en amont du gène FSCN1. ② Un segment de méthylation long de 17431 pb hyperméthylé est situé dans le gène CHRNA6. ③ Un segment hyperméthylé de 51 pb est également situé dans le gène CDH7. ④ La région autour du site d'initiation de la transcription (TSS) du gène CD9 est identifiée comme hypométhylée dans les CSC. ⑤ Et enfin, une région de 1000 pb hypométhylée est localisée à 8833 pb en aval du gène PRKAR1B. Ces différences de méthylation doivent maintenant être validées et corrélées à des changements d'expression de leur gènes associés.

4.2 ABSP : "ANALYSIS OF BISULFITE SEQUENCING PCR"



FONCTIONNEMENT EN BREF DE L'OUTIL ABSP

Pour chaque échantillon, l'application utilise en données d'entrée : les paramètres de l'expérience (nom de la séquence, condition, date, etc.), un fichier FASTA de la séquence d'ADN génomique de la région d'intérêt, et les deux fichiers ABIF (.ab1) de séquençage, un pour chaque direction. Une première analyse permet de contrôler la qualité des séquençages et de calculer les pourcentages de méthylation. Ces données sont ensuite récupérées pour une seconde analyse groupée de tous les échantillons. Cette dernière génère des graphiques de visualisation (*Lollipop style plots*) et compare les conditions entre elles par tests statistiques. Tous les résultats sont à la fois enregistrés dans les dossiers de l'application et compilés dans un rapport HTML pour parcourir tout le processus d'analyse.

APPORTS DE L'OUTIL ABSP

L'analyse des résultats de séquençage d'ADN convertis au bisulfite est longue et fastidieuse, c'est pourquoi de nombreux outils ont été développés. Ce nouvel outil ABSP présente de nombreux avantages par rapport aux outils existants : ① son processus d'analyse est complet, il propose en plus de l'estimation des pourcentages de méthylation, une visualisation des données et une analyse statistique pour déterminer les différences de méthylation entre plusieurs groupes d'échantillons, ② son utilisation est complètement automatisée pour l'utilisateur, de l'importation des données jusqu'à

la génération des graphiques et des comparaisons statistiques, ③ il prend en charge l'analyse des deux types de BSP, *direct-BSP* et *cloning-BSP*, et enfin ④ l'utilisation du langage de programmation libre R permet une très grande accessibilité, adaptabilité et évolutivité.

L'outil ABSP, ainsi que son manuel d'utilisation, exemples et données tests sont disponibles au téléchargement sur GitHub à l'adresse <https://github.com/ABSP-methylation-tool/ABSP>.

CONCLUSION

Grâce à une analyse globale de la méthylation de l'ADN, ce projet a permis d'amorcer l'identification de nouveaux acteurs intervenant dans le processus de dédifférenciation radio-induite de cellules non-CSC en CSC. Les différences de méthylation pourront ensuite être validées et corrélées à des changements transcriptomiques afin d'identifier des gènes potentiellement impliqués dans ce changement de phénotype. Enfin, l'étude plus poussée de ces gènes et de leur rôle dans la dédifférenciation radio-induite pourrait aboutir au développement de nouvelles solutions thérapeutiques pour prévenir l'enrichissement en CSC et ainsi radiosensibiliser les tumeurs.

Le programme ABSP apporte donc une nouvelle procédure d'analyse automatisée pour aider les biologistes à interpréter leurs résultats de BSP. En fournissant une solution clé en main pour analyser ces données, ABSP facilite ainsi l'accès à l'étude de la méthylation de l'ADN de régions d'intérêts. En effet, la technique de BSP étant très abordable, couplée à une analyse des résultats par ABSP, elle permet d'obtenir rapidement une réponse quant aux niveaux de méthylation d'une région spécifique de l'ADN.




Ainsi, ce travail a mis en lumière l'importance de la méthylation de l'ADN dans la plasticité du phénotype souche cancéreux et le potentiel d'amélioration des outils d'analyse.

TABLE OF CONTENTS

ABSTRACT	5
RÉSUMÉ	7
SYNTHÈSE	9
TABLE OF CONTENTS	15
LIST OF FIGURES	21
LIST OF TABLES	25
LIST OF CODE LISTINGS	27
ACRONYMS	29
I STATE OF THE ART	35
1 CANCER STEM CELLS: A CRUCIAL CHALLENGE IN THE FIGHT AGAINST CANCER	37
1.1 CANCER AND RESISTANCE TO THERAPIES	38
1.1.1 INDICATORS AND STATISTICS	38
1.1.2 THERAPIES AGAINST CANCER	41
1.1.3 RESISTANCE TO THERAPIES	44
1.1.4 CANCER STEM CELLS DRIVING THE CANCER RESISTANCE	45
📌 KEY POINTS	46
1.2 TUMOR DEVELOPMENT MODELS	46
1.2.1 STOCHASTIC (CLONAL EVOLUTION) MODEL	46
1.2.2 HIERARCHICAL (CSC) MODEL	47
1.2.3 PLASTICITY MODEL	48
1.2.4 UNIFIED MODEL	48
1.3 CONCEPT OF CANCER STEM CELLS	48
1.3.1 DISCOVERY AND DEFINITION	48
1.3.2 ORIGIN OF CANCER STEM CELLS	50
1.3.3 STEMNESS PROPERTIES	51
1.3.4 REGULATION OF STEMNESS	53
📌 KEY POINTS	58
1.4 RESISTANCE MECHANISMS OF CANCER STEM CELLS	58
1.5 THERAPEUTIC TARGETING OF CANCER STEM CELLS	59
1.5.1 CSC MARKERS	59
1.5.2 STRATEGY	60
1.5.3 APPROACHES	61
1.5.4 LIMITATIONS	62
📌 KEY POINTS	64
2 PLASTICITY AND DEDIFFERENTIATION INTO CANCER STEM CELLS	65

2.1	PHENOTYPIC PLASTICITY	65
2.1.1	DEFINITIONS	65
2.1.2	CSC PLASTICITY AND EMT	68
	📌 KEY POINTS	70
2.2	INDUCTION OF DEDIFFERENTIATION	70
2.2.1	SPONTANEOUS DEDIFFERENTIATION	72
2.2.2	THERAPY-INDUCED DEDIFFERENTIATION	73
2.2.3	FORCED REEXPRESSION OF PLURIPOTENCY FACTORS	75
	📌 KEY POINTS	76
2.3	REGULATION OF DEDIFFERENTIATION	76
2.3.1	HYPOXIA	76
2.3.2	EXTRACELLULAR VESICLES	78
2.3.3	INFLAMMATION AND SECRETED MOLECULES	79
2.3.4	EPIGENETICS	81
	📌 KEY POINTS	85
3	BREAST CANCER AND BREAST CANCER STEM CELLS	87
3.1	THE MAMMARY GLAND	87
3.1.1	DESCRIPTION	87
3.1.2	MAMMARY STEM CELLS	89
3.2	BREAST CANCER	90
3.2.1	INDICATORS AND STATISTICS	90
3.2.2	TUMORIGENESIS	92
3.2.3	TUMORAL PROGRESSION	94
3.2.4	HETEROGENEITY AND CLASSIFICATIONS	94
3.2.5	THERAPIES	99
3.3	BREAST CANCER STEM CELLS	101
3.3.1	MARKERS AND ISOLATION	101
3.3.2	FUNCTIONAL TESTS	103
	📌 KEY POINTS	104
4	DNA METHYLATION: PRINCIPLES AND METHODOLOGY	105
4.1	DNA METHYLATION	105
4.1.1	PRINCIPLE	105
4.1.2	ESTABLISHMENT, MAINTENANCE, AND REMOVAL	106
4.1.3	ROLE OF DNA METHYLATION	108
4.1.4	DNA METHYLATION AND EPIGENETIC REPROGRAMMING	109
4.1.5	DNA METHYLATION IN CANCER	110
4.1.6	DNA METHYLATION AND STEMNESS	111
4.1.7	EFFECTS OF RADIATIONS ON DNA METHYLATION	112
	📌 KEY POINTS	114
4.2	OVERVIEW OF METHODS TO STUDY DNA METHYLATION	114
4.2.1	BISULFITE-BASED ASSAYS	115
4.2.2	ENZYME DIGESTION-BASED ASSAYS	121
4.2.3	AFFINITY ENRICHMENT-BASED ASSAYS	122
	📌 KEY POINTS	123

4.3	BISULFITE SEQUENCING PCR (BSP)	123
4.4.1	EXPERIMENTAL PROCESS	124
	📌 KEY POINTS	130
4.4.2	ANALYSIS STRATEGY AND TOOLS	130
	📌 KEY POINTS	139
4.4.4	ARTIFACTS	139
	📌 KEY POINTS	147
	OBJECTIVES	149
II	MATERIAL AND METHODS	151
1	CELL CULTURE AND CANCER STEM CELLS POPULATION STUDY	153
1.1	CELL LINES AND CULTURE CONDITIONS	153
1.2	IRRADIATION	154
1.3	ALDH ACTIVITY STAINING	154
1.4	FLOW CYTOMETRY	154
1.5	FLUORESCENCE-ACTIVATED CELL SORTING (FACS)	155
1.6	SPHERE FORMING CAPACITY	155
2	RNA AND PROTEIN LEVELS	157
2.1	SIRNA TRANSFECTION	157
2.2	RNA LEVELS	157
2.3	PROTEIN LEVELS	159
3	DNA METHYLATION STUDY	163
3.1	REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS)	163
3.2	BISULFITE SEQUENCING PCR (BSP)	164
4	BIOINFORMATICS AND STATISTICS	167
4.1	DIFFERENTIALLY METHYLATED REGIONS (DMRs) IDENTIFICATION FROM REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS) DATA	167
4.1.1	TOOLS AND SOFTWARE	167
4.1.2	📌 PACKAGES	168
4.2	DEVELOPMENT OF ANALYSIS OF BISULFITE SEQUENCING PCR (ABSP)	169
4.2.1	TOOLS AND SOFTWARE	169
4.2.2	📌 PACKAGES	169
III	RESULTS	171
1	DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION	173
1.1	INVOLVEMENT OF EPIGENETIC-ASSOCIATED ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION	173
1.1.1	EXPRESSION OF EPIGENETIC-ASSOCIATED ENZYMES AFTER IRRADIATION	173
1.1.2	ROLE OF THE DNMT ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION	176
	📌 KEY POINTS	180

1.2	GLOBAL ANALYSIS OF DNA METHYLATION CHANGES DURING RADIO-INDUCED DEDIFFERENTIATION	180
1.2.1	STRATEGY AND SAMPLE PREPARATION	180
1.2.2	FIRST ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA	181
1.2.3	SECOND ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA	189
1.2.4	IDENTIFIED DIFFERENTIALLY METHYLATED REGIONS (DMRs)	197
	 KEY POINTS	204
2	DEVELOPMENT OF THE ABSP R TOOL FOR REGION-SPECIFIC CPG METHYLATION ANALYSIS	209
2.1	 PUBLICATION - "ABSP: AN AUTOMATED R TOOL TO EFFICIENTLY ANALYZE REGION-SPECIFIC CPG METHYLATION FROM BISULFITE SEQUENCING PCR"	209
2.2	ANALYSIS STRATEGY	218
2.2.1	DIRECT-BSP AND CLONING-BSP APPROACHES	218
2.2.2	GLOBAL WORKFLOW	219
2.3	AUTOMATIZATION AND INTERFACE	219
2.3.1	INSTALLATION AND OPENING	220
2.3.2	COMPONENTS OF ABSP	222
2.3.3	APP INTERFACE AND USABILITY	222
2.3.4	SCRIPT AUTOMATIZATION	227
2.3.5	ERROR HANDLING	228
2.3.6	MULTIPLE ANALYSES	231
	 KEY POINTS	235
IV	DISCUSSION AND PERSPECTIVES	237
	DISCUSSION AND PERSPECTIVES	239
1	IDENTIFICATION OF DNA METHYLATION CHANGES INVOLVED IN RADIO-INDUCED DEDIFFERENTIATION	239
1.1	GLOBAL EXPRESSION ANALYSIS OF EPIGENETIC ENZYMES	239
1.2	CONTRIBUTION OF DNMTs IN THE RADIO-INDUCED DEDIFFERENTIATION	240
1.3	ANALYSIS METHOD OF RRBS DATA TO IDENTIFY DIFFERENTIALLY METHYLATED REGIONS	241
1.4	DIFFERENTIALLY METHYLATED REGIONS AND ASSOCIATED GENES	244
2	THE ABSP TOOL: APPLICATION, LIMITATIONS, AND IMPROVEMENTS	247
V	BIBLIOGRAPHY	251
VI	APPENDIX 1	303
	PRIMER DESIGN FOR BISULFITE SEQUENCING PCR	305
1	PRIMER DESIGN SOFTWARE	305
2	GENERAL FEATURES	306
2.1	DNA TEMPLATE FOR PRIMER DESIGN	306
2.2	PRIMERS LENGTH	307
2.3	AMPLICON LENGTH	308
3	OTHER FEATURES	308
3.1	CPG SITES IN PRIMERS	308

3.2	CONVERTED CYTOSINES IN PRIMERS	309
3.3	PRIMERS GC CONTENT AND T_M DIFFERENCE	310
3.4	PRIMERS BASE COMPOSITION AND COMPLEMENTARITY	310
3.5	PRIMER SPECIFICITY AND UNINTENDED PCR PRODUCTS	310
3.6	AMPLICON BASE COMPOSITION: REPEATS, T_M DIFFERENCE AND CPG CONTENT	311
3.7	ADDITIONAL SEQUENCES	312
4	CONCLUSION	312
	📌 KEY POINTS	313
5	BIBLIOGRAPHY	313
VII	APPENDIX 2	317
	EXTRACELLULAR VESICLES AND DEDIFFERENTIATION INTO CANCER STEM CELLS	319
1	INTRODUCTION	319
1.1	EXTRACELLULAR VESICLES COMMUNICATION	319
1.2	EXTRACELLULAR VESICLES AND CANCER STEM CELLS	321
1.3	DEDIFFERENTIATION MEDIATED BY EXTRACELLULAR VESICLES	322
2	OBJECTIVE	323
3	MATERIAL AND METHODS	323
3.1	CELL LINES AND CULTURE CONDITIONS	323
3.2	EXTRACELLULAR VESICLES ISOLATION	323
3.3	EXTRACELLULAR VESICLES TREATMENT	324
3.4	RNA AND PROTEIN LEVELS	324
3.5	EXTRACELLULAR VESICLES UPTAKE	324
4	RESULTS	325
4.1	EXPERIMENTAL STRATEGY	325
4.2	CHARACTERIZATION OF THE RADIO-INDUCED EXTRACELLULAR VESICLES CONTENT	326
4.3	EFFECT OF RADIO-INDUCED EXTRACELLULAR VESICLES ON UNIRRADIATED CELLS	328
5	CONCLUSION	331
	📌 KEY POINTS	331
6	BIBLIOGRAPHY	331
VIII	APPENDIX 3	335
	ABSP USER GUIDE	337

LIST OF FIGURES

STATE OF THE ART

CANCER STEM CELLS: A CRUCIAL CHALLENGE IN THE FIGHT AGAINST CANCER

FIGURE 1	MAPS OF INCIDENCE AND MORTALITY RATES FOR ALL CANCERS IN 2020	39
FIGURE 2	ESTIMATED NEW CASES AND DEATHS BY CANCER SITES IN 2020	40
FIGURE 3	REPRESENTATION OF THE THREE TUMORAL DEVELOPMENT MODELS	47
FIGURE 4	UNIFIED MODEL OF TUMOR DEVELOPMENT AND DIVERSITY	49
FIGURE 5	THE ORIGIN OF CANCER STEM CELLS AT TUMOR INITIATION	50
FIGURE 6	DEFINING FEATURES OF CANCER STEM CELLS	51
FIGURE 7	POTENTIALITY OF NORMAL STEM CELLS	52
FIGURE 8	EPIGENETIC REGULATION OF CANCER STEM CELLS	56
FIGURE 9	RESISTANCE MECHANISMS OF CANCER STEM CELLS	60
FIGURE 10	THERAPEUTIC STRATEGY MODELS	61

PLASTICITY AND DEDIFFERENTIATION INTO CANCER STEM CELLS

FIGURE 11	PLASTICITY AND REPROGRAMMING EVENTS	66
FIGURE 12	NEW ADDITION TO HALLMARKS OF CANCER: UNLOCKING PHENOTYPIC PLASTICITY	67
FIGURE 13	CANCER CELL PLASTICITY: RELATION BETWEEN CSC PLASTICITY AND EMT	69
FIGURE 14	REGULATORS AND MECHANISMS INVOLVED IN THE DEDIFFERENTIATION OF NON-CSCs INTO CSCs	77

BREAST CANCER AND BREAST CANCER STEM CELLS

FIGURE 15	ANATOMY OF FEMALE MAMMARY GLAND	88
FIGURE 16	STRUCTURE AND CELL TYPE COMPOSITION OF THE MAMMARY GLAND	89
FIGURE 17	MAMMARY EPITHELIAL DIFFERENTIATION HIERARCHY	90
FIGURE 18	ESTIMATED NEW CASES AND DEATHS BY CANCER SITES FOR WOMEN IN 2020	91
FIGURE 19	BREAST CANCER PROGRESSION	95
FIGURE 20	BREAST CANCER MAIN SUBGROUPS CHARACTERISTICS	99
FIGURE 21	SCHEME OF CONVENTIONAL THERAPEUTIC CARE TO TREAT NON-METASTATIC BREAST CANCER	100

DNA METHYLATION: PRINCIPLES AND METHODOLOGY

FIGURE 22	METHYLATION OF CYTOSINE IN DNA	106
FIGURE 23	REPRESENTATION OF DNA METHYLATION STATES WITH ENZYMES INVOLVED	107
FIGURE 24	BISULFITE SEQUENCING PCR (BSP) EXPERIMENTAL PRINCIPLE	124
FIGURE 25	BISULFITE-MEDIATED DEAMINATION OF CYTOSINE	125
FIGURE 26	SCHEMATIC REPRESENTATION OF TOUCHDOWN AND NESTED PCR PROTOCOLS	127
FIGURE 27	AUTOMATED SANGER SEQUENCING PRINCIPLE	129
FIGURE 28	COMPARISON OF SEQUENCING TRACE DATA BETWEEN CLONING-BSP AND DIRECT-BSP APPROACHES	131
FIGURE 29	COMPARISON OF FEATURES BETWEEN THE THREE TOOLS ANALYZING CLONING-BSP DATA: QUMA, BIQ ANALYZER, AND BISMA	132
FIGURE 30	METHOD OF METHYLATION LEVEL CALCULATION BASED ON SEQUENCING CHROMATOGRAM FILE FOR DIRECT-BSP	134
FIGURE 31	THE EPIGENETIC SEQUENCING METHYLATION ANALYSIS SOFTWARE (ESME) TO EVALUATE METHYLATION FROM SEQUENCING TRACE FILES	136
FIGURE 32	VISUALIZATION OF METHYLATION DATA USING THE METHYLATION PLOTTER WEB TOOL	138
FIGURE 33	MELTING DIFFERENCES BETWEEN UNMETHYLATED AND METHYLATED BISULFITE CONVERTED DNA	144

MATERIAL AND METHODS

FIGURE 34	REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS) WORKFLOW	164
-----------	---	-----

RESULTS**DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION**

FIGURE 35	EXPRESSION OF THE EPIGENETIC-ASSOCIATED ENZYMES FOLLOWING IRRADIATION	175
FIGURE 36	INHIBITION OF DNMT1 AND DNMT3B BY SIRNA INTERFERENCE	177
FIGURE 37	EFFECT OF DNMT1 AND DNMT3B INHIBITION ON THE DEDIFFERENTIATION OF NON-CSCs INTO CSCs	178
FIGURE 38	EXPERIMENTAL RADIO-INDUCTION OF DEDIFFERENTIATION AND SUBPOPULATION ISOLATION FOR RRBS ANALYSIS	181
FIGURE 39	SCHEMATIC REPRESENTATION OF THE METHYLATION PROFILE OF A DMR POTENTIALLY INVOLVED IN DEDIFFERENTIATION	183
FIGURE 40	RRBS ANALYSES DIFFERENCES BETWEEN THE FIRST AND THE SECOND ONE	184
FIGURE 41	METHYLATION PROFILES OF DMRs ASSOCIATED WITH EPHB2, RUNX3, MAPK13, TGFB3, AND BMP2 GENES AND EXPRESSION CORRELATION	188

FIGURE 42	RRBS ANALYSIS WORKFLOW AND DATA PROCESSING	192
FIGURE 43	METHYLATION SEGMENTATION SUMMARY OF THE CSC 1 SAMPLE	195
FIGURE 44	VENN DIAGRAMS OF DMR SORTING	196
FIGURE 45	FILTERING DMRs BASED ON THEIR METHYLATION PROFILES INDICATING POTENTIAL INVOLVEMENT IN DEDIFFERENTIATION	201
FIGURE 46	METHYLATION PROFILES OF DIFFERENTIALLY METHYLATED REGIONS (DMRs) ASSOCIATED WITH FSCN1, CHRNA6, CDH7, CD9, AND PRKAR1B GENES	207

DEVELOPMENT OF THE ABSP R TOOL FOR REGION-SPECIFIC CPG METHYLATION ANALYSIS

FIGURE 47	ANALYSIS STRATEGY DIFFERENCES BETWEEN THE TWO BISULFITE SEQUENCING PCR APPROACHES, DIRECT-BSP AND CLONING-BSP	218
FIGURE 48	GENERAL WORKFLOW OF ABSP	220
FIGURE 49	DETAILED WORKFLOW OF ABSP	221
FIGURE 50	ABSP APP INTERFACE	227
FIGURE 51	REPORTS GENERATED BY THE INDIVIDUAL AND GROUPED ANALYSES	228
FIGURE 52	MULTIPLE WAYS TO LAUNCH THE ABSP ANALYSES	234

DISCUSSION AND PERSPECTIVES

BIBLIOGRAPHY

APPENDIX 1

PRIMER DESIGN FOR BISULFITE SEQUENCING PCR

FIGURE A1.1	PRIMER DESIGN GUIDELINES FOR BISULFITE SEQUENCING PCR (BSP)	309
-------------	---	-----

APPENDIX 2

EXTRACELLULAR VESICLES AND DEDIFFERENTIATION INTO CANCER STEM CELLS

FIGURE A2.1	EXTRACELLULAR VESICLES COMMUNICATION: BIOGENESIS AND UPTAKE	320
FIGURE A2.2	EXPERIMENTAL STRATEGY	326
FIGURE A2.3	CHARACTERIZATION OF THE EXTRACELLULAR VESICLES CARGO AFTER RADIOTHERAPY TREATMENT	327
FIGURE A2.4	EFFECT OF RADIO-INDUCED EXTRACELLULAR VESICLES ON UNIRRADIATED CELLS	330

APPENDIX 3

LIST OF TABLES

STATE OF THE ART

PLASTICITY AND DEDIFFERENTIATION INTO CANCER STEM CELLS

TABLE 1	MAIN STUDIES REPORTING THE DEDIFFERENTIATION OF NON-CSCs INTO CSCs	72
TABLE 2	HISTOLOGICAL GRADING SYSTEM OF BREAST CANCER	96
TABLE 3	CLINICAL STAGES OF BREAST CANCER BASED ON THE TNM CLASSIFICATION	97

DNA METHYLATION: PRINCIPLES AND METHODOLOGY

TABLE 4	LIST OF TECHNIQUES TO EVALUATE DNA METHYLATION	117
---------	--	-----

MATERIAL AND METHODS

CELL CULTURE AND CANCER STEM CELLS POPULATION STUDY

TABLE 5	CHARACTERISTICS OF THE SUM159PT BREAST CANCER CELL LINE	153
TABLE 6	COMPONENTS OF THE SUM159PT CELL LINE CULTURE MEDIUM	154

RNA AND PROTEIN LEVELS

TABLE 7	LIST OF siRNAs	157
TABLE 8	LIST OF SPECIFIC PRIMERS FOR RT-QPCR	158
TABLE 9	LYSIS BUFFER COMPOSITION FOR TOTAL PROTEIN EXTRACTION	159
TABLE 10	HYPOTONIC AND HYPERTONIC BUFFERS COMPOSITION FOR CYTOPLASMIC AND NUCLEAR PROTEIN EXTRACTION	160
TABLE 11	LIST OF ANTIBODIES USED FOR WESTERN BLOT	161

DNA METHYLATION STUDY

TABLE 12	PCR REACTION COMPOSITION FOR BISULFITE SEQUENCING PCR	165
TABLE 13	PCR BUFFER COMPOSITION FOR BISULFITE SEQUENCING PCR	165
TABLE 14	PRIMERS USED FOR BISULFITE SEQUENCING PCR	166

RESULTS

DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION

TABLE 15	SUMMARY STATISTICS OF CPG NUMBERS, COVERAGE, AND METHYLATION PERCENTAGES IN ALL SAMPLES THROUGHOUT THE ANALYSIS OF RRBS DATA	194
TABLE 16	PATHWAYS ENRICHMENT IN GENES ASSOCIATED WITH IDENTIFIED DMRs	200
TABLE 17	LIST OF FILTERED DIFFERENTIALLY METHYLATED REGIONS	204

DEVELOPMENT OF THE ABSP R TOOL FOR REGION-SPECIFIC CPG METHYLATION ANALYSIS

DISCUSSION AND PERSPECTIVES

BIBLIOGRAPHY

APPENDIX 1

PRIMER DESIGN FOR BISULFITE SEQUENCING PCR

TABLE A1.1	SOFTWARE THAT AID IN THE PROCESS OF DESIGNING PRIMERS FOR BSP	306
-------------------	---	-----

APPENDIX 2

EXTRACELLULAR VESICLES AND DEDIFFERENTIATION INTO CANCER STEM CELLS

TABLE A2.1	CHARACTERISTICS OF BREAST CANCER CELL LINES	323
TABLE A2.2	COMPONENTS OF BREAST CANCER CELL LINES CULTURE MEDIA	324
TABLE A2.3	LIST OF SPECIFIC PRIMERS FOR RT-QPCR	325
TABLE A2.4	LIST OF ANTIBODIES USED FOR WESTERN BLOT	325

APPENDIX 3

LIST OF CODE LISTINGS

RESULTS

DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION

CODE LISTING 1	GENERATING CPG METHYLATION PERCENTAGES FROM READ COUNTS.	193
CODE LISTING 2	GENERATING METHYLATION PROFILE SEGMENTS IN EACH SAMPLE	194
CODE LISTING 3	IDENTIFYING DMRs FROM THE 1,000 BP TILES SUBSET	197
CODE LISTING 4	ASSOCIATING DMRs WITH THEIR RELATED DATA	199

DEVELOPMENT OF THE ABSP R TOOL FOR REGION-SPECIFIC CPG METHYLATION ANALYSIS

CODE LISTING 5	APP SCRIPT: RENDERING INDIVIDUAL, GROUPED, AND MULTIPLE ANALYSES	231
CODE LISTING 6	PARAMETERS IN YAML HEADERS OF INDIVIDUAL AND GROUPED ANALYSES SCRIPTS	232
CODE LISTING 7	READING PARAMETERS IN INDIVIDUAL AND GROUPED ANALYSES SCRIPTS	233
CODE LISTING 8	HANDLING INCORRECT GENOMIC COORDINATES ERROR IN THE INDIVIDUAL ANALYSIS SCRIPT	234

ACRONYMS

Symbols | [A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [Y](#) | [Z](#)

Symbols

5caC 5-carboxylcytosine [108](#)

5fC 5-formylcytosine [108](#)

5hmC 5-hydroxymethylcytosine [106](#), [108](#)

5mC 5-methylcytosine [105](#), [106](#), [108](#), [114](#), [122](#), [123](#), [125](#), [126](#)

A

ABC ATP-binding cassette [59](#), [102](#), [245](#)

ABIF Applied Biosystems, inc. format [129](#), [166](#)

ABSP Analysis of Bisulfite Sequencing PCR [149](#), [166](#), [169](#), [218–220](#), [222](#), [227](#), [231](#), [235](#), [248](#), [249](#)

ADC antibody-drug conjugate [62](#), [64](#)

ADH atypical ductal hyperplasia [94](#), [95](#)

AH atypical hyperplasia [94](#)

ALDH aldehyde dehydrogenase [59](#), [60](#), [62](#), [102](#), [104](#), [154](#), [177](#), [240](#), [328](#)

ALH atypical lobular hyperplasia [94](#), [95](#)

AML acute myeloid leukemia [49](#), [62](#), [83](#), [110](#)

AR androgen receptor [97](#), [98](#)

B

BER base excision repair [59](#), [108](#), [110](#)

BiTE bispecific T cell engager [62](#), [64](#)

BMP bone morphogenetic protein [84](#), [110](#), [183](#)

BRCA breast cancer gene [94](#), [110](#)

BS-seq bisulfite sequencing [120](#)

BSP bisulfite sequencing PCR [23](#), [26](#), [115](#), [118](#), [123–128](#), [130](#), [137](#), [139](#), [143](#), [146](#), [147](#), [149](#), [201](#), [231](#), [235](#), [246](#), [247](#), [249](#), [305](#), [306](#), [309](#), [312](#)

C

CAF cancer-associated fibroblast [58](#), [79](#), [322](#)

CAR-T chimeric antigen receptor T [62](#), [64](#)

CCL C-C motif chemokine ligand [80](#), [81](#)

CCND1 cyclin D1 [93](#)

CD cluster of differentiation [59](#), [78](#), [83](#), [101](#), [326](#)

CDH7 cadherin 7 [201](#), [204](#), [207](#), [245](#), [246](#)

CGI CpG island [106](#), [108](#), [109](#), [112](#), [114](#), [120](#), [131](#), [142](#), [183](#), [184](#), [188](#), [189](#), [193–195](#), [198](#), [200](#), [204](#), [207](#), [241–243](#), [245](#), [246](#), [312](#)

CHRNA6 cholinergic receptor nicotinic alpha 6 subunit [201](#), [204](#), [207](#), [245](#)

CIS carcinoma *in situ* 94, 95
COBRA combined bisulfite restriction analysis 115, 121
cODC C-terminal degron of murine ornithine decarboxylase 60, 102
CpG cytosine-phosphate-guanine 106–108, 110, 114, 128, 130–147, 163, 166, 180, 181, 185, 188, 189, 192–194, 198, 200, 201, 204, 207, 218, 219, 227, 235, 240–245, 248, 249, 307, 308, 310–313
CSC cancer stem cell 45, 47, 49–51, 58, 65, 70, 101, 111, 149, 154, 155, 180, 241, 321
CXCL C-X-C motif chemokine ligand 80, 81
CXCR C-X-C motif chemokine receptor 80

D

DAMP damage associated molecular pattern 74
DCIS ductal carcinoma *in situ* 94, 95
ddNTP dideoxynucleotide triphosphate 119, 129, 133, 135
DEAB diethylaminobenzaldehyde 154, 176
DMR differentially methylated region 23, 112, 181–185, 188, 189, 192, 194–198, 200, 201, 204, 207, 242–246
DNMT DNA methyltransferase 56, 57, 62, 82, 83, 106–109, 111–114, 149, 175–177, 179, 180, 239–241, 247
dNTP deoxynucleotide triphosphate 129, 165

E

EBRT external beam radiotherapy 41
ECM extracellular matrix 57, 58
EMT epithelial-to-mesenchymal transition 68–70, 75, 81, 244, 245, 321–323
EMT-TF EMT-inducing transcription factor 68, 69
ER estrogen receptor 90, 96–98, 100, 104
ESC embryonic stem cell 52, 53, 55–57, 109–111, 113, 239
ESME epigenetic sequencing methylation analysis software 135–137, 139, 247, 248
EV extracellular vesicle 78, 79, 81, 85, 319–326, 328, 330, 331
EZH2 enhancer of zeste homolog 2 57, 84, 175, 176, 180, 239

F

FACS fluorescence-activated cell sorting 70, 72, 155, 175, 176, 179, 180, 241
FOXP1 forkhead box P1 244, 245
FSCN1 fascin actin-bundling protein 1 198, 200, 201, 204, 207, 245, 246

G

GFP green fluorescence protein 73
Gy gray 41, 74, 75

H

H2AK119Ub1 histone H2A lysine 119 monoubiquitination 56, 84
H3K27me3 histone H3 lysine 27 trimethylation 56, 82–84, 108, 175, 247
H3K4me3 histone H3 lysine 4 trimethylation 56, 83, 108, 247
HDAC histone deacetylase 62, 75, 76, 83, 84, 108, 112

HELP HpaII tiny fragment enrichment by ligation-mediated PCR 121, 122
HER2 human epidermal growth factor receptor 2 93, 96–100, 104
Hh Hedgehog 54, 55, 246
HIF hypoxia-inducible factor 58, 59, 73, 74, 76–78, 85
HMGB1 high-mobility group box 1 74
HSC hematopoietic stem cell 49, 52, 75
HTML hypertext markup language 222, 227, 228, 231

I

IARC international agency for research on cancer 39, 40, 91
IC invasive carcinoma 94, 95
iCSC induced cancer stem cell 180, 181, 241
IDC invasive ductal carcinoma 94, 95
IGV Integrative Genomics Viewer 306
IHC immunohistochemistry 96
IL interleukin 57, 74, 80, 81
ILC invasive lobular carcinoma 94, 95
inon-CSC irradiated cancer non-stem cell 180, 181, 241
iPS induced pluripotent stem 53, 57, 63, 66, 76, 82, 110

J

JAK Janus-activated kinase 55, 80, 85

K

KDM lysine demethylase 108, 175, 176, 180, 239
KLF4 kruppel like factor 4 53, 59, 66, 68, 75, 78, 110, 326

L

LC3-II light chain 3-II 78
LCIS lobular carcinoma *in situ* 94, 95
lncRNA long non-coding RNA 57, 79, 81, 82, 322

M

MAPK mitogen-activated protein kinase 93
MaSC mammary stem cell 52, 89, 90, 102
MBD methyl-CpG-binding domain 108, 111, 122
MDRE methylation-dependent restriction enzyme 121
MeDIP methylated DNA immunoprecipitation 122
MET mesenchymal-to-epithelial transition 68
MIP methylation-independent PCR 142, 143, 305–308, 313
MIRA methylated CpG island recovery assay 122
miRNA micro RNA 57, 79, 81, 321, 322
MREBS methylation-sensitive restriction enzyme bisulfite sequencing 120

MS-HRM methylation-sensitive high resolution melting 143, 144

MSP methylation-specific PCR 115, 121, 123, 124, 306

MSRE methylation sensitive restriction enzyme 120–122

N

NGS next-generation sequencing 114, 120, 122, 123, 129, 181, 248

NICD Notch intracellular domain 54, 61

non-CSC cancer non-stem cell 48, 49, 63, 65, 111, 149, 180, 241, 321

O

OCT4 octamer-binding transcription factor 4 53, 55, 57, 59, 66, 73, 75, 76, 78, 80, 109–111, 245, 326, 328

OXPHOS oxidative phosphorylation 59

P

PCR polymerase chain reaction 114, 115, 118, 119, 121–124, 126–131, 133–135, 139–147, 158, 164–166, 169, 192, 201, 218, 219, 305, 307, 308, 310–313, 324, 326

PGRN pluripotency gene regulatory network 53

PI3K phosphatidylinositol 3-kinase 80, 93

POU5F1 pit-oct-unc class 5 homeobox 1 53, 109

PR progesterone receptor 96–98, 100, 104

PRC polycomb repressive complex 56, 57, 82, 84, 108, 175

PRKAR1B protein kinase CAMP-dependent type I regulatory subunit beta 201, 204, 207, 245, 246

PTM post-translational modification 55, 56, 83, 85, 175, 247

R

RIN2 Ras and Rab interactor 2 244

RMD R markdown 222, 227, 228

ROS reactive oxygen species 42, 43, 59, 321

RRBS reduced representation bisulfite sequencing 120, 124, 163, 167, 168, 180, 181, 184, 185, 189, 192, 241–244, 246–248

RYR2 ryanodine receptor 2 244

S

SAHA suberoylanilide hydroxamic acid 75

SAM S-adenosylmethionine 106

SC stem cell 48–51

SERM selective estrogen receptor modulators 100

SFC sphere-forming capacity 60, 103, 104, 155, 176, 178, 179, 240

siRNA small interfering RNA 176, 177, 179, 180, 240, 241

SMO smoothed 55, 62

SNAI Snail family transcriptional repressor 68, 245, 326, 328, 331

SNPs single nucleotide polymorphisms 118, 134, 135

SOX sex determining region Y-box 53, 55, 57, 59, 66, 68, 73, 75, 76, 78, 83, 84, 110, 245, 328

SSC somatic stem cell 52, 53

STAT signal transducer and activator of transcription 55, 74, 80, 85, 183, 322

T

T_a annealing temperature 127, 128, 144–147

T_m melting temperature 126, 127, 142–145, 147, 305, 306, 308, 310, 311, 313

TAM tumor-associated macrophage 45, 58

TD-PCR touchdown PCR 128

TDLU terminal duct lobular unit 88, 94

TET ten-eleven translocation 56, 82, 85, 106, 108, 110, 114, 175, 176, 180, 239

TGF- β tumor growth factor β 55, 57, 79, 80, 183, 185, 245, 322

TIC tumor-initiating cell 49

TNBC triple-negative breast cancer 62, 97, 101, 104

TNF α tumor necrosis factor α 80, 85

Treg regulatory T cell 45

TSA trichostatin A 75

TSS transcription start site 112, 184, 185, 189, 193, 195–198, 200, 204, 242, 246

TWIST Twist family BHLH transcription factor 68, 80, 110, 326

U

UI user interface 222, 227

V

VPA valproic acid 75

W

WGBS whole-genome bisulfite sequencing 112, 120, 121, 124, 241, 242, 248

WHO world health organization 39, 40, 91

Y

YAML yet another markup language 227

Z

ZEB zinc finger E-box binding homeobox 68, 69, 80–83, 326

STATE OF THE ART

1

CANCER STEM CELLS: A CRUCIAL CHALLENGE IN THE FIGHT AGAINST CANCER

☰ CHAPTER CONTENTS

1.1	CANCER AND RESISTANCE TO THERAPIES	38
1.1.1	INDICATORS AND STATISTICS	38
1.1.2	THERAPIES AGAINST CANCER	41
1.1.3	RESISTANCE TO THERAPIES	44
1.1.4	CANCER STEM CELLS DRIVING THE CANCER RESISTANCE	45
	📌 KEY POINTS	46
1.2	TUMOR DEVELOPMENT MODELS	46
1.2.1	STOCHASTIC (CLONAL EVOLUTION) MODEL	46
1.2.2	HIERARCHICAL (CSC) MODEL	47
1.2.3	PLASTICITY MODEL	48
1.2.4	UNIFIED MODEL	48
1.3	CONCEPT OF CANCER STEM CELLS	48
1.3.1	DISCOVERY AND DEFINITION	48
1.3.2	ORIGIN OF CANCER STEM CELLS	50
1.3.3	STEMNESS PROPERTIES	51
1.3.4	REGULATION OF STEMNESS	53
	📌 KEY POINTS	58
1.4	RESISTANCE MECHANISMS OF CANCER STEM CELLS	58
1.5	THERAPEUTIC TARGETING OF CANCER STEM CELLS	59
1.5.1	CSC MARKERS	59
1.5.2	STRATEGY	60
1.5.3	APPROACHES	61
1.5.4	LIMITATIONS	62
	📌 KEY POINTS	64

⤴ [Back to Table of Contents](#)

1.1 CANCER AND RESISTANCE TO THERAPIES

This section aims to provide a brief overview of anti-cancer therapies and causes of therapeutic resistance, to understand the challenge of targeting cancer stem cells to diminish cancer recurrence.

1.1.1 INDICATORS AND STATISTICS

WORLDWIDE

In 2020, 19.3 million new cases of cancer and 10.0 million deaths have been reported worldwide by the GLOBOSCAN project (Figure 1) (Sung *et al.*, 2021). The 5-year prevalence, meaning the number of people alive within 5 years after a cancer diagnosis, is estimated to be 50.6 million people.

The three most commonly diagnosed cancers are female breast cancer (11.7%), lung cancer (11.4%), and colorectal cancer (10.0%), and the ones leading to the most deaths are lung cancer (18%), colorectal cancer (9.4%) and liver cancer (8.3%) (Figure 2).

Breast cancer is the most diagnosed cancer worldwide, representing 2.26 million new cases in 2020, 11.7% of all cancer diagnosed for both sexes, and 24.5% of all cancers diagnosed in women. It caused 685,000 deaths worldwide in 2020, 6.9% of all cancer deaths for both sexes, and 15.5% of all cancer deaths. Breast cancer is therefore the first cause of cancer mortality for women. The 5-year prevalence is estimated at 7.79 million, which is 17.7% of all cancers for both sexes, and 33.7% of all cancers for women, meaning that one-third of the women alive within 5 years after the cancer diagnosis, have been diagnosed with breast cancer (Sung *et al.*, 2021). Further details, specific to breast cancer, are given in chapter 3 “Breast cancer and breast cancer stem cells” at page 87.

It is estimated that one-quarter of men and one-fifth of women worldwide develop cancer during their lifetime and that one-eighth of men and one-eleventh of women worldwide die from cancer. The incidence is predicted to increase to 30.2 million new cases in 2040, almost 50% more than in 2020 (Global Cancer Observatory website).

IN FRANCE

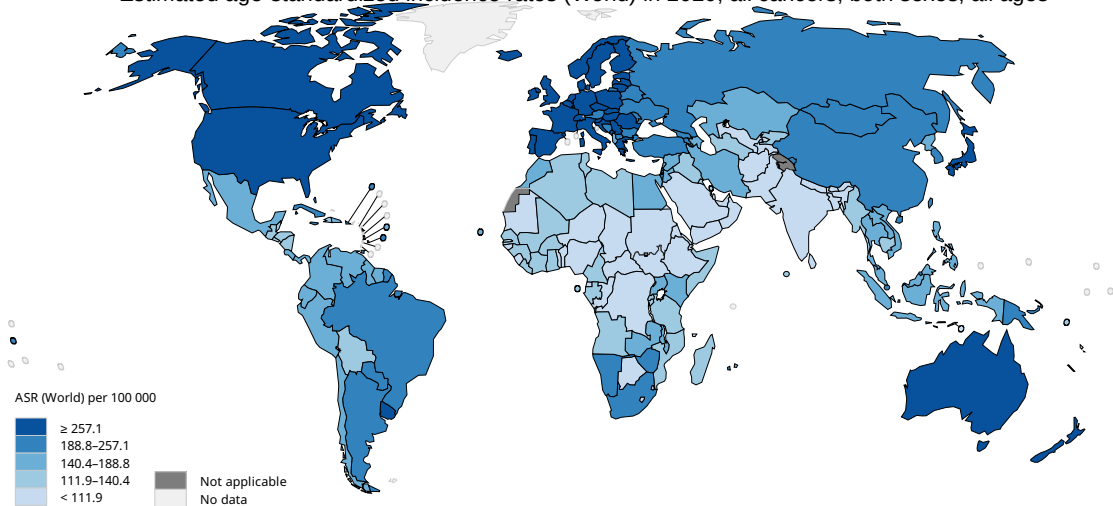
In France, 382,000 new cases of cancer and 157,400 deaths have been reported in 2018. The prevalence for 2017 is estimated at 3.8 million people alive who have been diagnosed with cancer in France.

The three most commonly diagnosed cancers in France are female breast cancer (15.3%), male prostate cancer (13.2%), and lung cancer (8.1%), and the ones leading to the most deaths are the lung cancer (21%), colorectal cancer (10.9%) and female breast cancer (7.7%), in 2018 (The French National Cancer Institute (INCa) website; Panorama des cancers en France, 2021).

With 58,500 new cases and 12,100 deaths in France in 2018, the breast cancer is also the

A

Estimated age-standardized incidence rates (World) in 2020, all cancers, both sexes, all ages



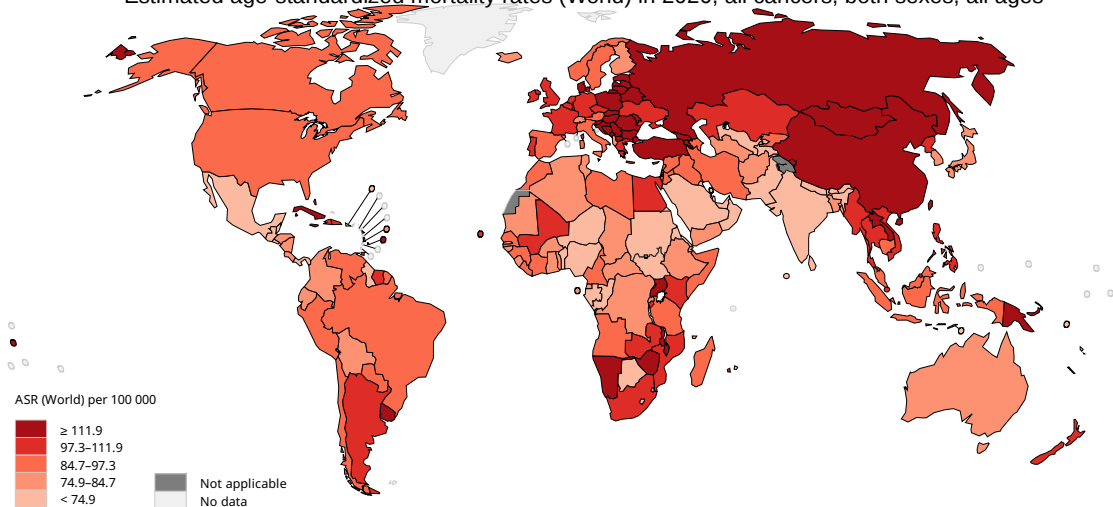
All rights reserved. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization / International Agency for Research on Cancer concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate borderlines for which there may not yet be full agreement.

Data source: GLOBOCAN 2020
Graph production: IARC
(<http://gco.iarc.fr/today>)
World Health Organization



B

Estimated age-standardized mortality rates (World) in 2020, all cancers, both sexes, all ages



All rights reserved. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization / International Agency for Research on Cancer concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate borderlines for which there may not yet be full agreement.

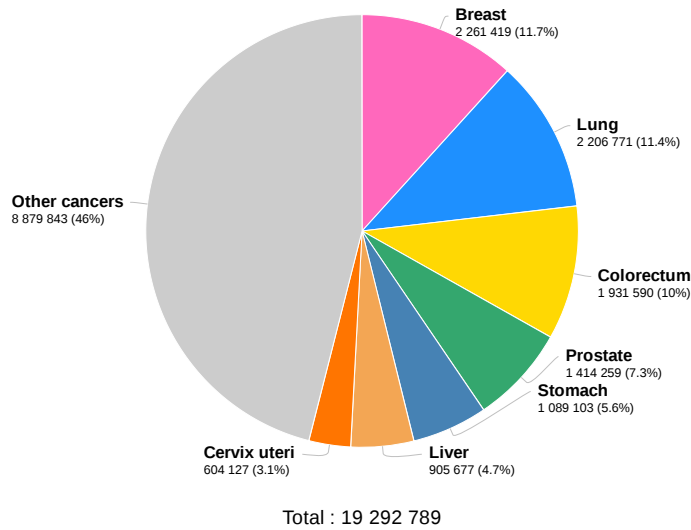
Data source: GLOBOCAN 2020
Graph production: IARC
(<http://gco.iarc.fr/today>)
World Health Organization



FIGURE 1 MAPS OF INCIDENCE AND MORTALITY RATES FOR ALL CANCERS IN 2020. **A.** Map of estimated incidence rates in 2020, for all cancers, both sexes, and all ages (ASR = age-standardized rate). **B.** Map of estimated mortality rates in 2020, for all cancers, both sexes, and all ages (ASR = age-standardized rate). Graph and data from GLOBOSCAN 2020, international agency for research on cancer (IARC), world health organization (WHO), available at the [Cancer Today - Global Cancer Observatory website](http://gco.iarc.fr/today).

A

Estimated number of new cases in 2020, worldwide, both sexes, all ages

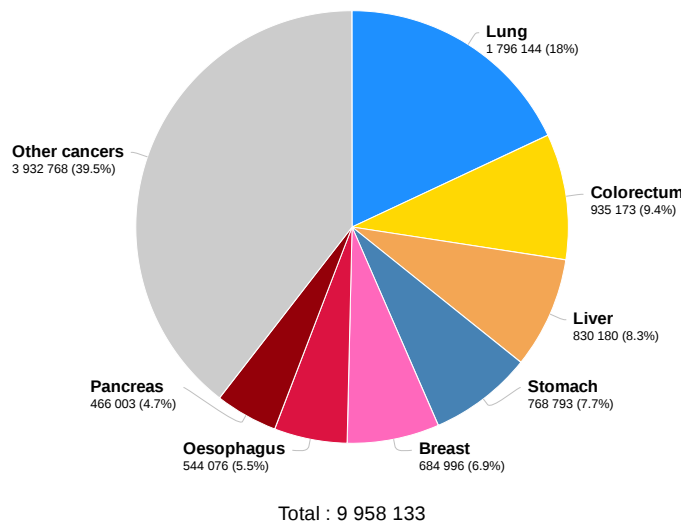


Data source: Globocan 2020
Graph production: Global Cancer Observatory (<http://gco.iarc.fr>)

International Agency for Research on Cancer
World Health Organization

B

Estimated number of deaths in 2020, worldwide, both sexes, all ages



Data source: Globocan 2020
Graph production: Global Cancer Observatory (<http://gco.iarc.fr>)

International Agency for Research on Cancer
World Health Organization

FIGURE 2 ESTIMATED NEW CASES AND DEATHS BY CANCER SITES IN 2020. **A.** Estimated number of new cases (incidence) in 2020, worldwide, for both sexes and all ages. **B.** Estimated number of deaths (mortality) in 2020, worldwide, for both sexes and all ages. Graph and data from GLOBOSCAN 2020, IARC, WHO, available at the [Cancer Today - Global Cancer Observatory website](http://cancer-today.org).

most diagnosed cancer (15.3% for both sexes, 33% for women) and the first cause of cancer mortality for women (17.8%) in France ([The French National Cancer Institute \(INCa\) website](#); [Panorama des cancers en France, 2021](#)). Further details, specific to breast cancer, are given in chapter 3 “Breast cancer and breast cancer stem cells” at page 87.

1.1.2 THERAPIES AGAINST CANCER

Several therapeutic protocols exist to treat different types of cancers, including chemotherapy, radiation therapy (or radiotherapy), surgery, immunotherapy, targeted therapy, and hormone therapy ([National Cancer Institute \(US\) website](#)). These treatments can be used in combination to be as effective as possible. The neoadjuvant therapy is the one administered before the main treatment, intending to reduce the tumor size beforehand to maximize the main therapy’s effectiveness. The main treatment is therefore called adjuvant therapy. The choice of treatments depends on the cancer site and specific biomarkers found in the cancer cells. Here is presented a brief overview of therapy principles, with a focus on radiotherapy. Details on therapies used to treat breast cancers are presented in section 3.2.5 “Therapies” at page 99.

RADIOTHERAPY

Radiation therapy, or radiotherapy, is a local treatment using **ionizing radiations** as a physical agent to destroy cancer cells. The radiation used is called ionizing because it induces the formation of ions (electrically charged particles) by detaching orbital electrons from atoms. Ionizing radiations can have an electromagnetic form, such as high-energy photons, or a particulate form, such as electrons, protons, neutrons, ions, or α particles. These particles have sufficient kinetic energy to ionize atoms by collision as they penetrate matter ([Dunne-Daly, 1999](#); [Gieringer et al., 2011](#); [Winiecki, 2020](#)). Gray (Gy) is the basic unit of radiation absorbed dose, corresponding to the amount of energy absorbed per unit mass.

Brachytherapy (also known as internal radiotherapy or *Curiethérapie*) and radioimmunotherapy utilize radionuclides (radioisotopes) as source of radiation in the form of α (two protons and two neutrons) particles and β particles (positrons or electrons) or γ -rays. In **brachytherapy**, the therapeutic radionuclides is placed on the skin surface (superficial) or is injected into the body (intracavitary, intraluminal, intravascular or interstitial) to be directly in contact with the tumor, therefore reducing normal tissue exposure to radiations ([Chargari et al., 2019](#); [Tanderup et al., 2017](#)). **Radioimmunotherapy** uses radiolabeled antibodies consisting of an antibody specific to a tumor-associated antigen coupled with a radionuclide, to deliver the ionizing radiation to the target cancer cells ([Larson et al., 2015](#); [Pouget et al., 2011](#)).

External beam radiotherapy (EBRT) is the most common form of radiotherapy, it utilizes beams of particles such as protons, neutrons, or electrons created by linear accelerators (linac), producing high-energy photons known as X-rays, that are directed at the tumour from outside the

body. Hence, different types of external beam radiotherapy exist, including electron beam therapy (electron beams) and proton therapy (proton beams) (Gieringer *et al.*, 2011; Hawley, 2013; Winiiecki, 2020).

Fractionation is the administration of repeated daily low doses of radiation over an extended period of time within a course of treatment, so that a high dose is given to the tumor, while ideally sparing normal tissue. A fractionated treatment is biologically less effective than a single radiation dose but minimize the damage applied to normal surrounding tissues (Falk, 2009; Hogle, 2006).

The biologic effects of fractionated radiation therapy on a tissue, either normal or malignant, relies on four principles, known as the **four “Rs” of radiobiology** (Falk, 2009; Hogle, 2006; Pajonk *et al.*, 2010; Withers, 1975):

- **Repair** (or recovery): the sublethal damage applied to cells is repaired depending on the cell type and radiation dose-rate.
- **Redistribution** (or reassortment): cells have different sensitivity to radiations depending on their cell cycle phase. Maximum effect from radiation occur just before and during cell division (G2 and M phases). Thus, the surviving tumoral population after radiation exposure is non-uniformly distributed through the cell cycle and reach the mitotic phase as the next dose is given, which increases the treatment effectiveness.
- **Repopulation** (or regeneration): cells respond to depopulation by regeneration through cell divisions between fractions, but the repopulation rate differ between normal and cancer cells. Normal cells can go through cell division to repopulate the tissue and prevent further damage, while tumor cells accumulate radiation effects and the one that succeeded in that succeeded in dividing within the course of the treatment are usually incapable of surviving. Fractionation reduces normal tissue damage while optimizing the killing of tumor cells.
- **Reoxygenation**: as well-oxygenated cells are more radiosensitive than cells in a low oxygen environment (hypoxia) as within the tumor mass, the fractionated treatment allows the tumor to shrink, causing cells to become more oxygenated and thus less radioresistant.

Additionally, the intrinsic radiosensitivity of cells as well as the reactivation of anti-tumor immune response also modulate the biological effect of radiotherapy (Boustani *et al.*, 2019).

After penetrating the tissue, radiations begin to ionize surrounding molecules, destroying them by breaking down chemical bounds between atoms.

Radiations generate **reactive oxygen species (ROS)** (superoxide anion $O_2^{\cdot-}$, hydroxyl radicals $\cdot OH$ and hydrogen peroxide H_2O_2) by the radiolysis of water from the microenvironment and they are highly reactive entities toxic for cells. Moreover, radiations induce the production of endogenous ROS in mitochondria. Excessive intracellular ROS levels disrupt the redox system balance causing oxidative stress by reacting with biological molecules (proteins, lipids, nucleic acids) (Kim *et al.*, 2019b; Renschler, 2004).

The **DNA molecules are damaged** by direct ionization or by the oxidative stress caused by ROS levels increase, inducing base oxidation, apurinic or apyrimidinic sites, single-strand breaks and double-strand breaks. So, in response to radiations, DNA damage repair mechanisms are activated to maintain DNA fidelity and the cell cycle is arrested to allow time for repair. Errors in the DNA repair lead to accumulation of mutations, while persistence of too much DNA damage induces cell death mechanisms such as apoptosis or mitotic catastrophe. Cells can also enter into a senescent state, corresponding to a stable irreversible state outside of the cell cycle (G0 phase) (Baskar *et al.*, 2012; Kim *et al.*, 2019b).

In addition, the **radiation treatment have non-targeted effects on unirradiated cells**, including genomic instability perpetuation in the descendants of irradiated cells, bystander effects (regional effects, e.g. release of signaling molecules or extracellular vesicles affecting unirradiated cells) and abscopal effects (distant effects, e.g. radiation-induced immune response affecting unirradiated metastases) (Bright and Kadhim, 2018; Wang *et al.*, 2015a).

OTHER THERAPIES

Chemotherapy is the use of drugs as a systemic treatment to kill cancer cells based on their highly proliferative phenotype. This way, chemotherapy aims to slow or stop tumor growth to reduce its size and associated symptoms. Several types of cytotoxic agents exist, targeting various processes associated with cell cycle checkpoints, DNA replication, and DNA damage, to prevent cell division and ultimately result in cell death by apoptosis. As examples: anti-metabolites are structural analogs of purines or pyrimidines and interfere with the biosynthesis of nucleic acids by substituting for normal bases; alkylating agents induce the alkylation of DNA bases and thereby damage the DNA; cross-linking agents bind covalently between two strands of DNA to create DNA cross-linking, impairing processes such as replication or transcription; anti-tubulin agents impair the mitotic spindle formation and block cell division; and topoisomerase inhibitors suppress the topoisomerase activity that loosens the DNA supercoiling during replication and transcription and thus, impair these essential processes (Nussbaumer *et al.*, 2011; Parnell and Woll, 2005). As chemotherapeutic drugs are toxic to all dividing cells present in the organism, it causes side effects (e.g. hair loss, digestive system dysregulation) as this toxicity impacts normal cells as well, and it is, therefore, a major limit to the use of these drugs. Hence, the balance between high efficiency and reduced toxicity is found by modulating protocols, by adding recovery periods for instance.

Surgery, or resection, consists in removing the tumor or a part of it from the body by a surgical procedure. **Immunotherapy** aims to assist the immune system in the anti-tumor response, to overcome cancer cells immunosuppressive capacities. **Targeted therapy** consists in targeting specific proteins or molecular pathways that are necessary for tumor growth by using specific inhibitors such as small-molecule drugs or antibodies for example. **Hormone therapy** is used on cancers depending on hormones to grow, such as prostate or breast cancer. It aims to slow tumor growth by blocking

hormone-associated mechanisms or pathways.

1.1.3 RESISTANCE TO THERAPIES

The major limitation to therapy effectiveness is the cancer resistance to treatment. The **resistance to therapy** can be defined as a therapeutic failure, it happens when the cancer is no longer responding to the therapy and the tumor starts to grow again. Consequently, it causes the recurrence of tumors and ultimately, relapse or death. As the incidence of cancers increases, it raises the necessity to continuously improve treatments to fight against cancer (Hanahan and Weinberg, 2000; Longley and Johnston, 2005; Sporn, 1996).

A cancer therapy includes three factors: the therapy, the targeted cancer cell population, and the host environment. Therefore, the therapeutic resistance results from **a combination of factors, at several levels, varying over time**. Therapeutic resistance is not only limited to the intrinsic and acquired duality of resistance properties at the tumor cell and tumor microenvironment levels as often described (Alfarouk *et al.*, 2015; Vasan *et al.*, 2019). The intrinsic resistance refers to the innate capacities of the tumor to resist a treatment before the therapy. The acquired resistance describes the gradual acquisition by the tumor of new ways to resist the therapy, resulting in a gradual reduction of the therapy effectiveness (Wang *et al.*, 2019b). Hence, the resistance is multi-level and multi-factorial as it is the result of several overlapping factors including, the pharmacological properties of the therapy, intrinsic and acquired resistant phenotypes of cancer cells, and extrinsic environmental factors (Alfarouk *et al.*, 2015). Accordingly, the cancer resistance to therapy is not only a matter of cellular resistant phenotypes. All of the tumor cells do not have to be qualified as resistant for the tumor to resist therapy and start regrowing, only the survival and expansion of some resistant cells may be responsible for it, in a particular environment and with specific treatment properties. Here is presented an overview of resistance to therapies occurring at different levels.

First, for systemic treatments using drugs, such as chemotherapy or targeted therapy, at the **macroscopic level or systemic level**, the pharmacokinetics: absorption, distribution, metabolism, and excretion (ADME) of the drug by the organism, are involved in the resistance as these parameters impact the effectiveness of the drug to reach tumor cells, reduce their proliferation and kill them (Izar *et al.*, 2013; O'Connor, 2007).

Additionally, at a **mesoscopic level or regional level**, the tumor-host interactions and the physics of the tumor site have also a role in drug resistance. The vascular morphology and the intratumoral blood flow have a direct consequence on the drug intratumoral delivery and the potential stagnation sites of drugs (Alfarouk *et al.*, 2015; Minchinton and Tannock, 2006; Salnikov *et al.*, 2003). And, by alteration of the intratumoral blood flow, cancer cells can reduce their exposure to drugs to create a favorable hypoxic environment (low oxygen levels) for their proliferation (Minchinton and Tannock, 2006; Vasan *et al.*, 2019). Also, the drug will not reach the tumor population when this one is located in specific protected areas, such as the central nervous system protected by the blood-brain

barrier (Osswald *et al.*, 2016; Puhalla *et al.*, 2015). In the case of radiotherapy treatment, the radiation of cancer cells, as well as non-cancer cells, affects intercellular communication, immune system's anti-tumoral activities, and thereby the tumor microenvironment. The radiation treatment triggers the relapse of extracellular factors, which has immunosuppressive effects, by suppressing immune cell activation for example, and pro-tumorigenic effects, through the enhancement of DNA repair capacity for example (Ashrafizadeh *et al.*, 2020). Immune cells within the tumor microenvironment, such as tumor-associated macrophages (TAMs) or regulatory T cells (Tregs), are also particularly affecting the response to immunotherapy treatments. For example, the loss of T cell function or the diminution of T cell recognition are mechanisms that can lead to acquired resistance to immunotherapeutic drugs (Sharma *et al.*, 2017).

Lastly, at the **microscopic level or local level**, resistance to therapy comes from the tumor cells themselves, through innate or acquired properties. These resistance properties are thereby caused by cellular level changes, differing from normal cells, but also differing among the cancer cells, resulting in a heterogeneous population in terms of therapy response. Indeed, tumor cells have differences in cellular morphologies, gene expression, signaling pathway activities, epigenetic patterns, motility, metabolism, proliferation, and metastatic potential, as a result of genetic variations and environmental factors (Dagogo-Jack and Shaw, 2018; Haider *et al.*, 2020). The genetic variety is due to genomic instability, meaning an increased mutation rate, and chromosomal instability and rearrangements, meaning the gain or loss of the whole chromosome (aneuploidy) or structural aberrations. Genomic instability is mainly caused by a high division rate with DNA replication defects and important DNA damage with impaired repair mechanisms (Lengauer *et al.*, 1998; Sansregret *et al.*, 2018; Stephens *et al.*, 2011). This instability gives rise to gene mutations and gene amplification affecting the cell phenotype, which varies among the tumor population.

By conferring advantages regarding survival and proliferation, some genomic alterations are selectively preserved and lead to the expansion of specific competitive clones (McGranahan and Swanton, 2017). This evolution model is called the clonal evolution model of tumor cell populations and was first described by Nowell (1976). As therapies induce genomic instability in cancer cells, they enhance the selection of clones with resistant properties. Thereby, under selective therapeutic pressure can emerge new adaptive responses in tumor cells driving toward resistance (Vasan *et al.*, 2019).

1.1.4 CANCER STEM CELLS DRIVING THE CANCER RESISTANCE

Within the tumor heterogeneity, one particular subpopulation has been described and characterized as a major contributor to cancer resistance: the so called cancer stem cells (CSCs). This subpopulation have been reported to contribute to therapeutic resistance in several cancer models, such as leukemia (Viale *et al.*, 2009), glioblastoma (Eramo *et al.*, 2006), pancreatic cancer (Hermann *et al.*, 2007) and breast cancer (Chuthapisith *et al.*, 2010). The concept of CSC will be further

developed within the next sections.

Interestingly, the pre-existent CSCs, by having higher resistance properties, are selected upon treatment. Due to their intrinsic properties (tumor initiation, self-renewal abilities and differentiation potential), CSCs can give rise to a heterogeneous population of cancer cells (Najafi *et al.*, 2018). This CSC-driven tumor progression contributes to therapeutic failure and cancer relapse (Frank *et al.*, 2010; Phi *et al.*, 2018). Indeed, many studies have reported that a high proportion of CSCs is correlated with poor prognosis (Ginestier *et al.*, 2007; Mertins, 2014).

Altogether, it raises the necessity to target both the tumor bulk and the CSC population to reduce the tumor resistance and disease progression in patients. Therefore, finding new therapeutic solutions to target CSCs is now a major challenge in the field of cancer resistance research.

KEY POINTS

- Cancers represent 19.3 million new cases and 10 million death in 2020 worldwide. Several therapies exist to treat cancers, such as surgery, chemotherapy, radiotherapy, and immunotherapy.
- The major limitation to therapy effectiveness is resistance to treatments. It drives therapeutic failure as tumor cells no longer responds to the treatment and starts to progress again. The mortality caused by cancer is widely attributed to the therapeutic resistance and to the formation of new tumors in distant sites.
- A response to therapy depend on three compounds: the therapy, the cancer cells, and the host environment. Thereby, the therapeutic resistance is the result of several factors including, the pharmacological properties of the therapy, intrinsic and acquired phenotypes of cancer cells, and extrinsic environmental factors, operating at several levels, systemic, regional, and local levels.
- A subpopulation of cells, cancer stem cells, drives the cancer resistance to therapies. By having more efficient resistance mechanisms, differentiation, and self-renewal abilities, these cells can regenerate the bulk of cancer cells and regrow the tumor upon treatment. Hence, it raises the necessity to study them to find new insights and new therapeutic solutions to fight against cancer resistance.

↗ [Back to Table of Contents](#)

1.2 TUMOR DEVELOPMENT MODELS

1.2.1 STOCHASTIC (CLONAL EVOLUTION) MODEL

Historically, the tumor development was described as following a clonal evolution model, or stochastic model, in which the tumor cell heterogeneity is explained by various clones originating from a succession of different mutations or epigenetic alterations (Figure 3) (Nowell, 1976; Rich, 2016; Torres *et al.*, 2007). This evolution model follows a Darwinian process, whereby the clones having a selective advantage becomes dominant ones, participating in the tumor progression. The resistance to treatments is therefore explained by the selection of the more resistant clones under

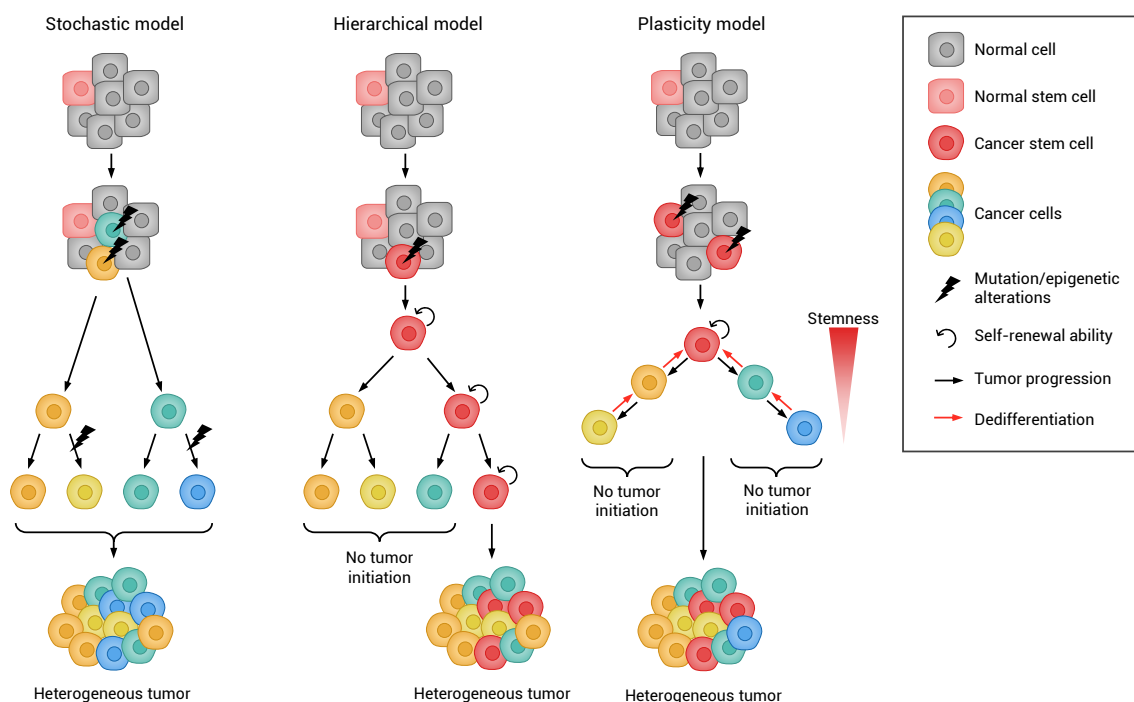


FIGURE 3 REPRESENTATION OF THE THREE DEVELOPMENT MODELS. The **stochastic model** (or clonal evolution model) postulates that genomic or epigenetic alterations lead to tumor formation as clones are equipotent, they all possess the tumorigenic potential to initiate the tumor formation and are responsible for its heterogeneity. The **hierarchical model** (or CSC model) postulates that a subpopulation called cancer stem cell (CSC) holds alone the potential for tumor initiation and regeneration of a heterogeneous tumor population, through self-renewal and differentiation. The **plasticity model** postulates expand the hierarchical/CSC model by postulating that cancer cells lacking tumorigenic potential can dedifferentiate into CSC and acquire the associated stem-like properties. Figure adapted from [Dick \(2008\)](#); [Fulawka et al. \(2014\)](#); [Gimple et al. \(2019\)](#); [Rich \(2016\)](#); [Thomas et al. \(2019\)](#).

the therapeutic pressure. In this model, all the cells are equipotent, meaning all of them to have the potential to regenerate a tumor.

1.2.2 HIERARCHICAL (CSC) MODEL

The hierarchical model, or CSC model, proposes a hierarchic organization of tumor cells, with the CSC population at the top, similarly to the normal tissue organization ([Figure 3](#)) ([Bonnet and Dick, 1997](#); [Rich, 2016](#)). As this population is able to self-renew and differentiate to regenerate an heterogeneous population of cancer cells, these cells are therefore considered responsible for tumor progression and heterogeneity. Indeed, the CSCs alone can regenerate the entire tumor with all of primary tumor diversity ([Al-Hajj et al., 2003](#); [Dick, 2008](#)). Thereby, coupled with their intrinsic resistance mechanisms, these cells are designated as a major cause of resistance to therapies and metastasis formation ([Bandhavkar, 2016](#); [Najafi et al., 2019](#)).

The stochastic and hierarchical models are not mutually exclusive, both could co-exist within the tumor and explain its heterogeneity, especially since CSCs can be seen as a mix of clones

whereby the clonal selection can be applied (Shipitsin and Polyak, 2008).

1.2.3 PLASTICITY MODEL

The **plasticity model** extends the CSC model, by adding a reversibility of the stem-like state (Figure 3) (Michor and Polyak, 2010; Plaks *et al.*, 2015; Rich, 2016). Indeed, it has been described that differentiated cancer cells, or cancer non-stem cells (non-CSCs), can reacquire stem cells properties in several cancer models (Chaffer *et al.*, 2011; Debeb *et al.*, 2012; Lagadec *et al.*, 2012; Schwitalla *et al.*, 2013; Yang *et al.*, 2012). As cancer cells are more plastic than normal cells, this model describes the CSC phenotype as a transitory state, comparable with the EMT process. Intrinsic tumor factors or stimuli from the microenvironment can influence the shift between states of cancer cells and induce the reacquisition of stem cell characteristics. The CSCs differentiate into transit-amplifying cells, or progenitors (direct progeny of CSCs), mildly pluripotent, proliferative and lacking self-renewal ability, that can revert to a CSC state (Aponte and Caicedo, 2017; Nassar and Blanpain, 2016). The stemness hierarchy remains in this model as progenitor cells cannot regenerate the full tumor heterogeneity, but it is less rigid as cells can change state and move up in the hierarchy towards CSCs, instead of just going down towards differentiation. Further explanations on plasticity and dedifferentiation of cancer cells are given in chapter 2 “Plasticity and dedifferentiation into cancer stem cells” at page 65.

1.2.4 UNIFIED MODEL

These three models are not exclusive but complementary as they all participate in increasing the tumor diversity and they can be gathered in a unifying model, as proposed by Michor and Polyak (2010) (Figure 4). Indeed, inside the CSC compartment, CSCs can evolve by acquiring additional genetic mutations and can be clonally selected under environmental pressure, participating in cancer progression and resistance to therapies.

The accumulation of mutations in CSCs are perpetuated due to their self-renewal ability and transmitted through differentiation to transit-amplifying cells. These cells give rise to non-CSC that are also subjected to genetic alteration events, but they cannot persist as they are not able to self-renew unless they reacquire self-renewal abilities through dedifferentiation and rejoin the CSC pool (Michor and Polyak, 2010). This model supports the theory of the CSC phenotype as a transient state whereby accumulated mutations conferring a selective advantage persist and are spread in more differentiated cells (Figure 4).

1.3 CONCEPT OF CANCER STEM CELLS

1.3.1 DISCOVERY AND DEFINITION

In adult tissues, normal stem cells (SCs) are responsible for tissue maintenance and regeneration, thanks to their self-renewal ability and capacity to differentiate into multiple cell lineage required

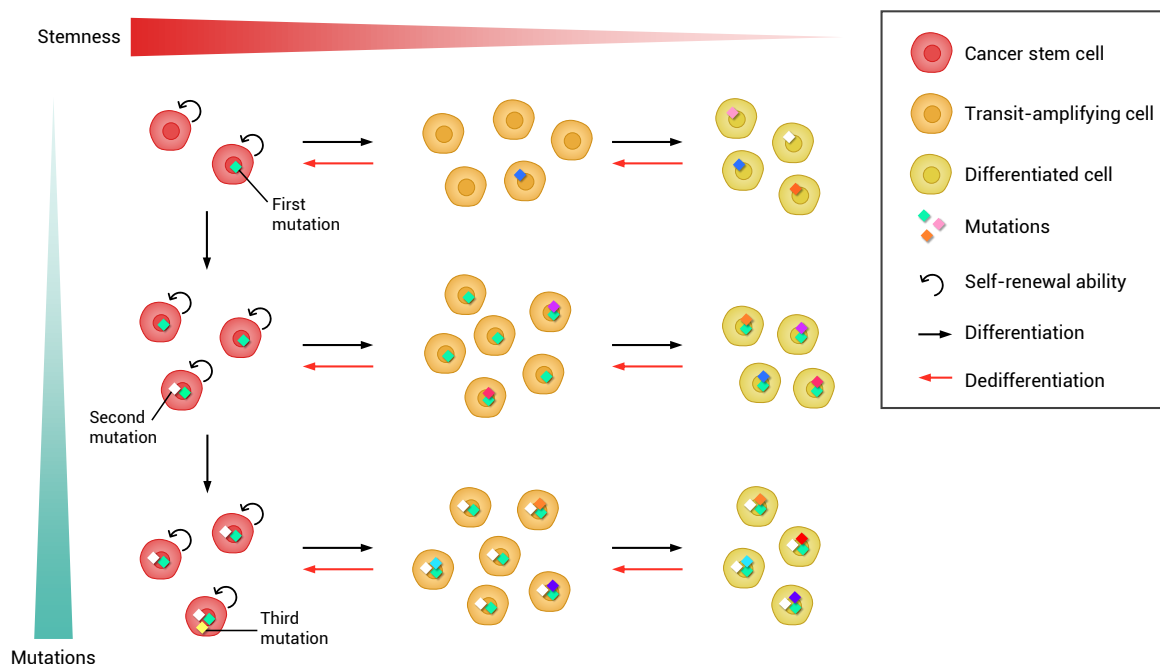


FIGURE 4 UNIFIED MODEL OF TUMOR DEVELOPMENT AND DIVERSITY. This model proposes to combine the three models of tumor development: stochastic (clonal evolution), hierarchical (CSC), and plasticity models. The CSC pool accumulates genetic mutations, participating in increasing the diversity of tumor cells and in the tumor progression. These mutations persist through the self-renewal of the CSC population and are transmitted to transit-amplifying cells through differentiation. These cells become non-CSC cells and also accumulate mutations that cannot persist unless the cell reacquires self-renewal abilities by rejoining the CSC population through dedifferentiation. Figure adapted from [Michor and Polyak \(2010\)](#).

for a specific organ ([Blanpain and Fuchs, 2014](#)). This implies a hierarchical organization of tissues, that has been particularly described in the hematopoietic tissue ([Eaves, 2015](#)). Based on this model, [Bonnet and Dick \(1997\)](#) applied this concept to human acute myeloid leukemia (AML). They already demonstrated that leukemia cells had different capacities to propagate leukemia when transplanted into immunodeficient mice ([Lapidot et al., 1994](#)). They found that the leukemia-initiating fraction of cancer cells possess specific cell surface markers, described as the $CD34^+/CD38^-$ cells. This population, called leukemia stem cells (LSCs), or CSCs, have similar properties of normal SCs. Indeed, they express genes similar to those expressed by hematopoietic stem cells (HSCs), they are able to self-renew and can regrow an heterogeneous cancer population by regeneration of different cell lineages. It was thereby established that similarly to normal tissues, tumors can be organized in a hierarchical model ([Bonnet and Dick, 1997](#)). Later on, the presence of CSCs was demonstrated in several other cancer models, including breast cancer ([Al-Hajj et al., 2003](#)), glioma ([Singh et al., 2004](#)), colorectal cancer ([O'Brien et al., 2007](#)) and pancreatic cancer ([Li et al., 2007a](#)).

By definition, cancer stem cells (CSCs) are cancer cells capable to produce more CSCs as well as to differentiate into cancer cells, enabling the regeneration of an heterogeneous tumor. That's why CSCs are also called tumor-initiating cells (TICs) ([Clarke et al., 2006b](#)), which could be confusing

regarding the cell of origin of the tumor.

1.3.2 ORIGIN OF CANCER STEM CELLS

As the CSCs are driving the tumor initiation and progression, it raises the question of their origin at the point of tumor initiation. Two hypothesis remains, either the CSCs come from normal differentiated cells or adult tissue SCs (Figure 5).

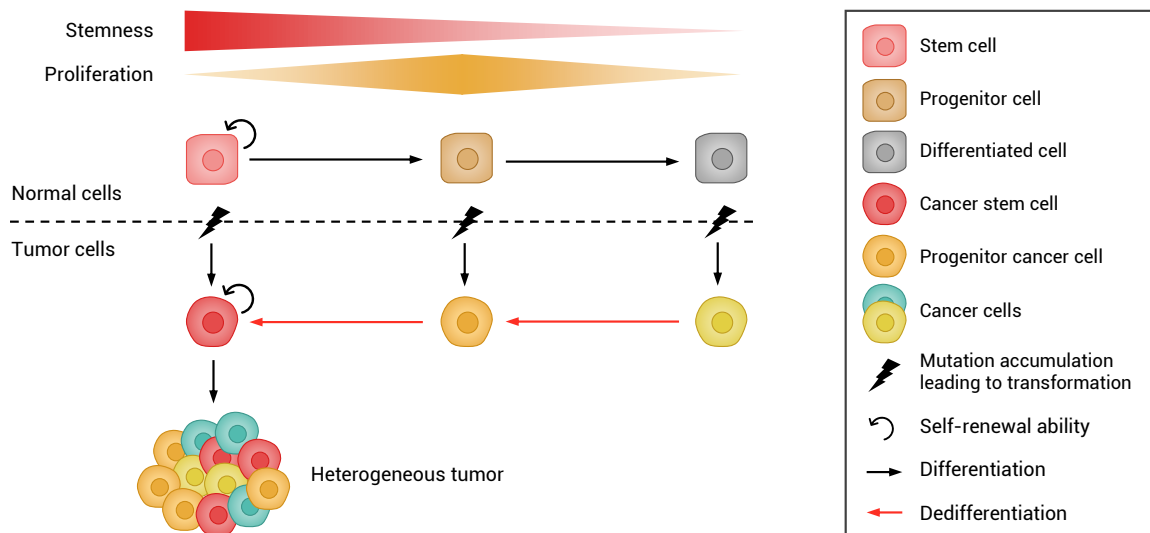


FIGURE 5 THE ORIGIN OF CANCER STEM CELLS AT TUMOR INITIATION. Adult stem cells, progenitors, or differentiated cells can be transformed by an accumulation of mutations and initiate the tumor development. For progenitors and differentiated cells to be the cell of origin of the tumor, they must reacquire stemness properties such as self-renewal through a dedifferentiation process, to generate the heterogeneity of tumor cells. The type and aggressiveness of the tumor vary depending on the cell of origin. Figure adapted from [Walcher et al. \(2020\)](#).

First, it was hypothesized that CSCs were coming from the transformation of an adult SC, in which oncogenes are overexpressed and tumor suppressors inactivated, promoting uncontrolled growth of the cells, and initiating the tumor development (details in section 3.2.2 “Tumorigenesis” at page 92) ([Hanahan and Weinberg, 2011](#); [Sutherland and Visvader, 2015](#)). As the SCs already have unlimited growth potential, only a few genetic changes would be required for their transformation, giving a tumor-initiating cell with stemness properties ([Walcher et al., 2020](#)). However, studies have demonstrated, by the lineage tracing of cancer cells, that tumors can also originate from differentiated cells or progenitor cells ([Blanpain, 2013](#); [Mu et al., 2015](#); [Oikawa, 2016](#); [Perekatt et al., 2018](#)). The plasticity of transformed cells is enough for their dedifferentiation through the re-acquisition of stemness characteristics, induced by the accumulation of mutations and/or environmental factors (Figure 5). Additionally, it has been shown that the cell of origin influence the type, aggressiveness, and prognosis of the tumor (details in section 3.2.4 “Heterogeneity and classifications” at page 94 for breast cancer) ([Blanpain, 2013](#); [Visvader, 2011](#)).

1.3.3 STEMNESS PROPERTIES

The normal SCs possess several distinct characteristics, also called stemness properties, which mainly consists in self-renewing while also generating differentiated cells. In cancers, by possessing self-renewal and differentiation abilities, CSCs are able to drive the growth of a heterogeneous tumor while maintaining their pool. Interestingly, this process happens during tumor progression, after treatments or while driving the formation of heterogeneous metastases at distant sites (Figure 6) (Lytle *et al.*, 2018).

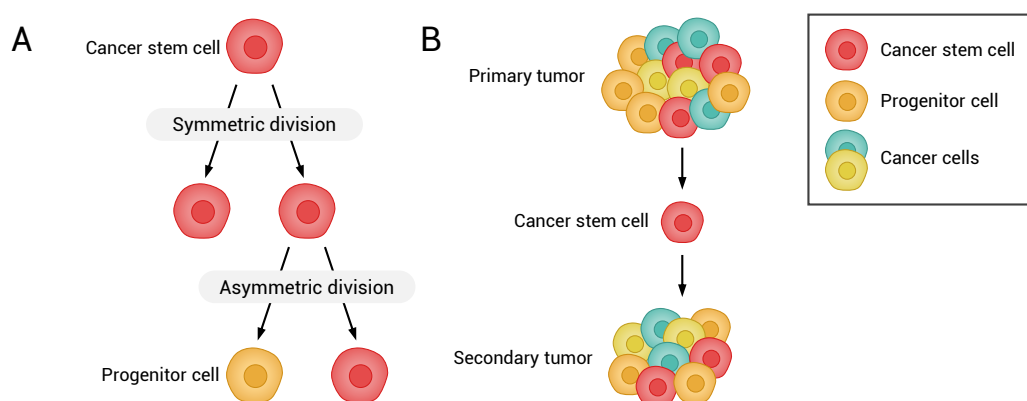


FIGURE 6 DEFINING FEATURES OF CANCER STEM CELLS. **A.** The self-renewal of CSCs is the ability to generate daughter cells with stemness characteristics. A CSC can undergo a symmetric division, giving two CSCs, or an asymmetric division, giving one CSC and one progenitor cell without stemness features and committed to a differentiation process. **B.** A CSC can regenerate a heterogeneous tumor population, which is possible due to their self-renewal and generation of differentiated cells abilities. Figure adapted from Fulawka *et al.* (2014).

SELF-RENEWAL

Self-renewal is the ability for cells to proliferate while maintaining a pool of cells with the same characteristics indefinitely. A cell division is described as symmetric when the cellular components are equally distributed and two identical daughter cells are produced. It's described as asymmetric when cellular components are unequally distributed, giving two different daughter cells, one possessing the same characteristics as the initial cell while the other is more differentiated (Figure 6A) (Fuchs and Chen, 2013).

The self-renewal ability along with the asymmetrical cell division are key features of normal stem cells for tissue homeostasis and regeneration over time (Urbán and Cheung, 2021). In tissues, the self-renewal of stem cells is highly regulated, as self-renewal impairment weakens tissue regeneration, while the over-activation of self-renewal can lead to cell transformation and tumor development. Additionally, a fine balance between symmetric and asymmetric divisions, often spatially regulated within the tissue, allow the maintenance of the stem cells pool (Fuchs and Chen, 2013).

Similar to normal stem cells, CSCs are capable of both symmetric and asymmetric division to

perpetuate themselves within a fast growing tumor (Lytle *et al.*, 2018; Schillert *et al.*, 2013).

POTENTIALITY

The potentiality is the capacity of unspecialized cells to differentiate into specialized cells along lineages. Different types of potentiality have been defined in normal stem cells depending on the range of lineages a cell can differentiate into: totipotency, pluripotency, multipotency, bipotency, and unipotency (Figure 7) (O'Connor and Crystal, 2006).

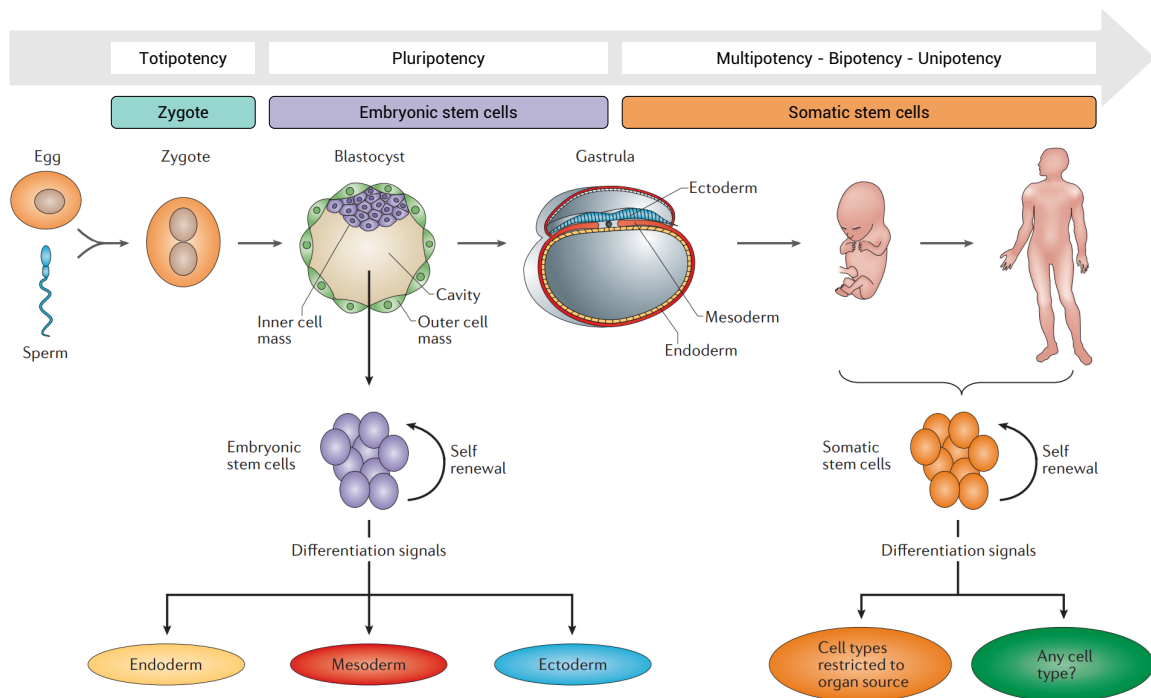


FIGURE 7 POTENTIALITY OF NORMAL STEM CELLS. After fertilization, the diploid zygote cell is totipotent and initiates a series of cell divisions to form the embryo. At the blastocyst stage, embryonic stem cells (ESCs) derived from the inner cell mass can differentiate into all cell types of the three primary germ layers, endoderm, mesoderm, and ectoderm, they are pluripotent. In the fetus and later in the adult, the somatic stem cells (SSCs) can differentiate into various cell types depending on their organ of origin, they are multipotent. When SSCs are able to differentiate into two lineages or one lineage, they are called bipotent or unipotent, respectively. Figure adapted from O'Connor and Crystal (2006).

First, the zygote formed after the fertilization is defined as totipotent as it corresponds to the original cell that will create all cells of the organism. Then, embryonic stem cells (ESCs) are pluripotent as they can differentiate into all cell types of the three primary germ layers, endoderm (gastrointestinal tract, lungs), mesoderm (bones, muscles, blood), and ectoderm (epidermis, nervous system). Finally, in the fetus and adult, the SSCs are more specialized and can differentiate into cell types of the tissue in which they reside, that's why they are called after their organ of residencies, such as mammary stem cells (MaSCs) or hematopoietic stem cells (HSCs). These cells can be either multipotent if they can differentiate into multiple lineages, bipotent if they can differentiate into

two lineages, or unipotent if they can differentiate into only one lineage (O'Connor and Crystal, 2006). The unipotency can be also associated with progenitors (direct daughter cells of stem cells) that are not fully differentiated and can still differentiate further in one lineage (Rodrigues *et al.*, 2019).

The potentiality of SSCs to differentiate into one or several lineages is essential for tissue homeostasis and regeneration. As tissues present different regeneration needs, the proliferation rate of stem cells depends on their tissue residency environment. In rapidly renewing tissues, such as the skin or the intestine, most stem cells proliferate continuously whereas, in low turnover tissues, such as muscles or the nervous system, stem cells are mostly in a quiescent state, meaning a non-proliferative state that can be reversed upon appropriate stimuli (Urbán and Cheung, 2021). Hence, not all SSCs are quiescent but quiescence is still a key feature of stem cells.

Similarly to SSCs, CSCs express pluripotency genes coding for transcription factors involved in the maintenance of the stem cell state.

1.3.4 REGULATION OF STEMNESS

PLURIPOTENCY FACTORS

Several genes have been reported as essential for the maintenance of the pluripotent state of normal stem cells, as well as in CSCs: **octamer-binding transcription factor 4 (OCT4)** (also named pit-oct-unc class 5 homeobox 1 (POU5F1)), **sex determining region Y-box (SOX) 2** and **NANOG** are the three main ones (Boyer *et al.*, 2005; Liu *et al.*, 2013a).

These pluripotency-associated genes code for master transcription factors regulating numerous gene programs, they are forming the core of the pluripotency gene regulatory network (PGRN), a cascade of regulatory events that maintain the self-renewal ability, the pluripotency state, and prevent the differentiation of the cell. The pluripotency factors act in complex, OCT4 and SOX2 form heterodimers while NANOG form homodimers, that bind to specific DNA consensus sites and have been found to co-occupy hundreds of potential regulatory elements in the genome, including other pluripotency factors genes such as kruppel like factor 4 (KLF4) and LIN28 (Li and Belmonte, 2017; Li and Belmonte, 2018). This OCT4–SOX2–NANOG occupancy has been found notably on enhancers associated with the regulation of self-renewal and differentiation. In addition, the OCT4/SOX2 complex also regulates the NANOG gene, which acts as a safeguard to repress the pro-differentiation signals (Rodda *et al.*, 2005). Other pluripotency factors, such as KLF4, C-MYC, and LIN28, have also been reported to be involved in the maintenance of a pluripotent state.

In ESCs, OCT4, SOX2, and NANOG function as differentiation repressors of the three germ layers (endoderm, mesoderm, ectoderm) lineages (Thomson *et al.*, 2011). The artificial reexpression of pluripotency factors leads to the reprogramming of somatic cells into an ES-like state by reestablishing their pluripotent state, these cells are called induced pluripotent stem (iPS) cells (Takahashi *et al.*, 2007; Takahashi and Yamanaka, 2006; Theunissen *et al.*, 2011; Yu *et al.*, 2007b).

In cancers, the overexpression of pluripotency factors is correlated with tumor aggressiveness and serves as CSC markers (Ben-Porath *et al.*, 2008; Chiou *et al.*, 2010; Chiou *et al.*, 2008).

SIGNALING PATHWAYS

A cross-talk of several signaling pathways, such as **Wnt/β-catenin**, **Notch** and **Hedgehog (Hh)** pathways, regulates the self-renewal and the maintenance of pluripotency, participating in the regulation of stemness by the microenvironment in both normal and cancer stem cells (Ajani *et al.*, 2015; Liu *et al.*, 2013b; Matsui, 2016; Yang *et al.*, 2020a).

The **Wnt/β-catenin pathway** can be conducted in two different ways, as a canonical pathway operating, through the β-catenin, involved in cell fate determination or as a non-canonical pathway, β-catenin independent, participating in cell movement and tissue polarity control. In absence of a Wnt signal, the β-catenin binds the GSK3-AXIN-APC destruction complex, composed of glycogen synthase kinase 3 (GSK3), Axin, and adenomatous polyposis coli (APC), which leads to its ubiquitination and subsequent degradation by the proteasome. In presence of the Wnt signal, Wnt ligands bind to the Frizzled family receptors and LRP5/LRP6 co-receptors causing a cascade of signaling interactions resulting in the disruption of the GSK3-AXIN-APC destruction complex. The β-catenin molecules translocate and accumulate into the nucleus, where they form a complex with T cell factor/lymphoid enhancer factor (TCF/LEF) family transcription factors and co-activators, leading to the transcription of target genes (Katoh and Katoh, 2007; Pohl *et al.*, 2017). Hence, the Wnt/β-catenin pathway contributes to stem cell maintenance, embryonic development, and tissue homeostasis (Mohammed *et al.*, 2016).

In cancer, the Wnt signaling has been associated with tumorigenesis, tumor progression, and therapy resistance, its over-activation is correlated with a poor prognosis and increased recurrence (Holland *et al.*, 2013; Katoh, 2017; Mohammed *et al.*, 2016; Yang *et al.*, 2015). Its activation promotes the self-renewal capacity of CSCs, in several cancer models such as leukemia and prostate cancer (Bisson and Prowse, 2009; Kawaguchi-Ihara *et al.*, 2008). It was also reported that Wnt signals orient the asymmetric division of CSCs by maintaining the stem-like state of one daughter cell while the other acquires differentiation features (Habib *et al.*, 2013)

The **Notch pathway** depends on physical interactions between adjacent cells to be activated. The ligand binding to the Notch receptor triggers the cleavage of the Notch extracellular domain (NECD) and the transmembrane domain, releasing the Notch intracellular domain (NICD). This NICD domain translocates to the nucleus and forms a transcriptional complex with DNA binding proteins and co-activators to induce the transcription of Notch target genes (Hori *et al.*, 2013). The Notch pathway contributes to the cell fate determination of normal stem cells during the development by maintaining a balance between pluripotency and differentiation (Bigas and Porcheri, 2018). In cancer, the Notch pathway is activated in CSCs and participates in the disease progression and resistance to treatments (Bolós *et al.*, 2008; Giuli *et al.*, 2021). In breast cancer, the radiotherapy treatment

induces the Notch signaling activation and the enrichment in CSCs (Lagadec *et al.*, 2013).

The **Hedgehog (Hh) pathway** is activated when the Hh ligands bind the patched (PTC) - smoothed (SMO) receptor complex, inducing a signaling cascade up to the activation of the glioma-associated oncogene (GLI) family of transcription factors, to trigger the expression of Hh target genes. Non-canonical signaling of Hh, ligation- and receptor-independent, have also been described (Carballo *et al.*, 2018). Target genes of the Hh signaling are involved in proliferation, differentiation, and survival, and includes the NANOG gene (Coni *et al.*, 2013). Thus, the Hh signaling regulates the self-renewal and pluripotency of stem cells and is important for embryogenesis, tissue homeostasis, and repair (Petrova and Joyner, 2014). In cancer, it also participates in the CSCs pool maintenance and in their resistance to treatment (Clement *et al.*, 2007; Liang *et al.*, 2021; Tanaka *et al.*, 2009).

Other signaling pathways implicated in the embryonic development and the tissue homeostasis, such as the **Janus-activated kinase (JAK) / signal transducer and activator of transcription (STAT) 3 pathway** or the **tumor growth factor β (TGF- β) pathway**, have also been reported to regulate stemness properties of CSCs by promoting their self-renewal (Jin, 2020; Sakaki-Yumoto *et al.*, 2013).

EPIGENETICS

Epigenetic mechanisms regulate the gene transcription within the cell, without altering the DNA sequence but only by changing the accessibility of genetic loci to transcriptional machinery through chromatin remodeling. Several types of epigenetic modifications are reported: nucleosome remodeling, histone variants, histone post-translational modifications (PTMs), DNA methylation, and non-coding RNAs.

Through these different mechanisms, the expression of genes linked to the CSC-associated pathways (e.g. Wnt, Notch, Hh) and pluripotency factors (e.g. OCT4, SOX2, NANOG) is finely regulated in both normal and cancer stem cells (Figure 8).

The **nucleosomes** are complexes of eight histone proteins: histones H2A, H2B, H3, and H4, each in dimers. The DNA molecule wraps around the nucleosome (147 bp) to form the basic unit of the chromatin (Clapier and Cairns, 2009). The accessibility of promoters to transcription factors depends on the conformation and the presence or absence of nucleosomes. Chromatin remodeling complexes, also called remodelers, can remove, slide, or restructure nucleosomes to modify the target site exposition to the transcription machinery (Jiang and Pugh, 2009; Morgan *et al.*, 2020). These remodelers, such as the switch/sucrose non-fermenting (SWI/SNF) complex, participate in the pluripotent state and self-renewal maintenance in ESCs and leukemia CSCs (Figure 8A) (Gao *et al.*, 2008; Shi *et al.*, 2013).

Incorporation of **histone variants** within the nucleosome core, like macroH2A1, H2A.Z, or H3.3, can also induce chromatin conformation changes by altering the structure and stability of nu-

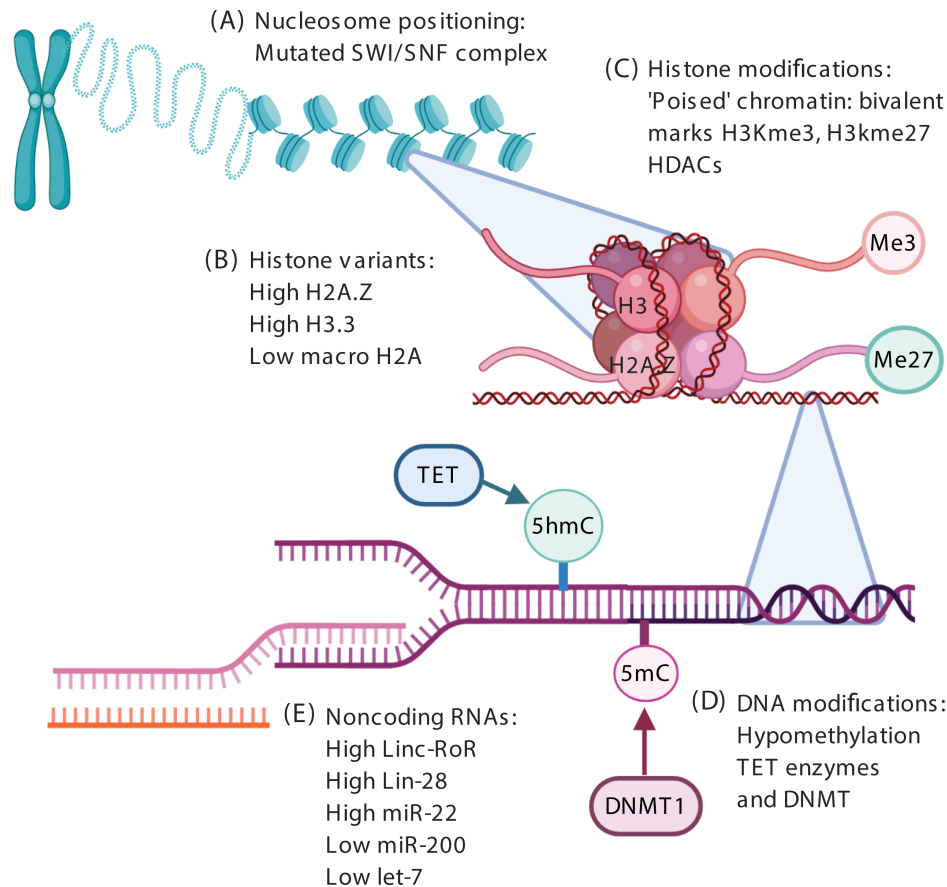


FIGURE 8 EPIGENETIC REGULATION OF CANCER STEM CELLS. Several epigenetic mechanism regulate the stemness-associated genes expression in CSCs. (A) Nucleosome remodelling and remodelers complexes. (B) Changes in histone variant deposition. (C) Histone post-translational modification and histone modifiers. (D) DNMT-mediated DNA methylation and TET-mediated DNA demethylation. (E) Changes in non-coding RNAs expression levels. Figure from French and Pauklin (2021).

cleosomes. The mobility of the core histones and histone variants have been associated with stem cell properties including pluripotency (Bošković *et al.*, 2014; Santenard and Torres-Padilla, 2009; Turinetto and Giachino, 2015). In CSCs, changes in histones variants macroH2A1 and H3.3 levels have been linked to the maintenance of their self-renewal capacity (Figure 8B) (Gallo *et al.*, 2015; Park *et al.*, 2016; Re *et al.*, 2018).

Histone post-translational modifications, such as methylation or acetylation of lysine residues at histones tails for example, affect gene expression by altering the chromatin state (euchromatin, active state, and heterochromatin, repressive state) and the ability of protein complexes to bind to target loci. For instance, the presence of histone H3 lysine 4 trimethylation (H3K4me3) at promoters is generally correlated to gene transcription activation, while the histone H3 lysine 27 trimethylation (H3K27me3) and histone H2A lysine 119 monoubiquitination (H2AK119Ub1), established by the polycomb repressive complex (PRC) 1 and 2, are associated with gene silencing. Distinct histone modification landscapes are found between differentiated cells and ESCs, pluripotent cells present

more euchromatin and less heterochromatin (Hawkins *et al.*, 2010). Histone modifying enzymes, acetylase, methylase, or demethylases, are therefore crucial for the regulation of pluripotency and cell identity (Figure 8C) (Boland *et al.*, 2014). The two repressive complexes PRC1 and PRC2 and in particular their respective catalytic subunits BMI1 and enhancer of zeste homolog 2 (EZH2) are required for pluripotency of both normal and cancer stem cells in several models (Cruz-Molina *et al.*, 2017; Liu *et al.*, 2006b; Proctor *et al.*, 2013; Vlerken *et al.*, 2013; Wen *et al.*, 2021).

DNA methylation participates to the regulation of genes programs associated with stemness features (Figure 8D). Further details on DNA methylation principles can be found in chapter 4 “DNA methylation: principles and methodology” at page 105. First, the promoter region of the NANOG pluripotency factor genes have been found hypomethylated in CSCs (Wang *et al.*, 2013c). Several studies have reported a stem cell DNA methylation signature in CSCs (Helou *et al.*, 2014; Lee *et al.*, 2015). Then, the up-regulation of DNA methyltransferases (DNMTs), responsible for establishment and maintenance of DNA methylation, is correlated with increased tumorigenic capacities of CSCs while demethylating agent treatment is correlated with a decrease of them (Liu *et al.*, 2014; Tian *et al.*, 2012; Tsai *et al.*, 2012). The treatment with DNMT inhibitors also leads to re-sensitization of ovarian cancer cells to chemotherapy agents (Wang *et al.*, 2014b). Lastly, the importance of TET-mediated demethylation for the pluripotency state maintenance of ESCs and its establishment in iPS cells was also reported (Costa *et al.*, 2013; Gao *et al.*, 2013; Olariu *et al.*, 2016).

Finally, several **long non-coding RNAs (lncRNAs)** and **micro RNAs (miRNAs)**, have also been reported to modulate pluripotency and self-renewal associated genes (Figure 8E). For instance, the expression of the lncRNA H19 is correlated with the expression of OCT4 and SOX2 pluripotency factors and with increased tumorigenic capacities of CSCs (Roy *et al.*, 2015). Additionally, H19 along with the let-7 miRNA family also regulates the LIN28 pluripotency factor expression, involved in the promotion of CSCs symmetric division and thereby self-renewal (Albino *et al.*, 2016; Lecerf *et al.*, 2020). Many other miRNAs have been described as involved in the regulation of CSCs stemness characteristics, such as the miR-200, miR-21, miR-22, and miR-183 families (Shimono *et al.*, 2015).

NICHE

CSCs reside in distinct regions within the tumor, called the CSCs niches, consisting in specialized environments that maintain their stemness properties while protecting them from therapeutic agents and immune system (Plaks *et al.*, 2015). This environment has specific physicochemical properties and is composed of stromal cells, immune cells, endothelial cells all releasing extracellular matrix (ECM) molecules, inflammatory factors and growth factors.

Different key components are forming the CSC niche. First, CSCs can interact with stromal cell through **cell-to-cell communication** (Melzer *et al.*, 2017). Either by the release of extracellular factors (e.g. cytokines such as interleukin (IL)-6 and IL-8 and growth factors such as the TGF- β) for a paracrine action or by juxtacrine interactions (e.g. Notch receptor, Ephrin receptors), the

stromal cells, such as cancer-associated fibroblasts (CAFs) or TAMs, can activate signaling pathways associated to stemness (Essex *et al.*, 2019; Lu *et al.*, 2014a; Nair *et al.*, 2017).

The spatial organization of the **extracellular matrix (ECM)** components (proteins, glycoproteins, proteoglycans, and polysaccharides) has a determinant role in the CSC niche. The ECM acts as a physical barrier blocking drugs, provides anchorage for CSCs, and its remodeling by matrix metalloproteinases facilitates the diffusion of extracellular factors and enhances stemness properties (Lu *et al.*, 2012). The anchorage provided by the ECM in the CSC niche is required for the maintenance of cell polarity, regulating the symmetrical and asymmetrical divisions of CSCs, thus the ECM is essential for the self-renewal and differentiation of stem cells (Yamashita and Fuller, 2008).

Another major feature of the CSC niche is its oxygen levels, determined by the tumor vascularization. CSCs can reside both in a hypoxic region (reduced oxygen levels) promoting their survival and self-renewal, or in a perivascular area, facilitating their dissemination through blood vessels (Plaks *et al.*, 2015). **Hypoxia** induces increased levels of the hypoxia-inducible factor (HIF) family of transcription factors, and HIF-1 α have been described as involved in the CSC pool maintenance, enrichment, and resistance to therapies (Carnero and Lleona, 2016; Rainho *et al.*, 2021).

Altogether, the stemness properties of CSCs are the result of an interconnected regulation, involving pluripotency transcription factors as master regulators, signaling pathways, epigenetic control of the chromatin and microenvironment interactions.

KEY POINTS

- A cancer stem cell (CSC) is defined as a tumor initiating cell, capable of regenerating a heterogeneous tumor, thanks to two main features: self-renewal capacity and pluripotency (Figure 6 and Figure 7).
- The CSC at tumor initiation can originate from either the transformation a normal stem cell that already possesses stemness features or from the transformation and dedifferentiation of differentiated normal cell (Figure 5).
- A fine regulation of self-renewal and pluripotency is established in CSCs, and involved a cross-talk of multi-levels mechanisms.
- The tumoral microenvironment within the CSC niche, epigenetic mechanisms (Figure 8), pluripotency-associated signaling pathways and pluripotency-associated transcription factors are altogether inter-operating the stemness properties of CSCs.

↗ [Back to Table of Contents](#)

1.4 RESISTANCE MECHANISMS OF CANCER STEM CELLS

Several interconnected mechanisms have been associated with CSC resistance to therapies, especially to chemotherapy and radiotherapy (Figure 9) (Garcia-Mayea *et al.*, 2019; Li *et al.*, 2021; Najafi *et al.*, 2019; Prieto-Vila *et al.*, 2017; Zhou *et al.*, 2021).

First, the previously described **stemness features**: self-renewal ability, associated signaling pathway and transcription factors, as well as the **CSC niche** promotes the maintenance of CSC upon treatments. For example, the Wnt/ β -catenin pathway up-regulates the production of **ATP-binding cassette (ABC) transporters** (Chau *et al.*, 2013; Milosevic *et al.*, 2020). These ABC transporters (e.g. ABCB1, ABCG2, ABCB5) are **drug efflux pumps** responsible for the elimination of cytotoxic agents including chemotherapeutic drugs and are highly expressed in CSCs (Begicevic and Falasca, 2017; DeGorter *et al.*, 2012). Additionally, a **protective autophagy machinery** can be over-activated in CSCs to minimize the cellular exposure to stress such as chemotherapeutic agents or radiotherapy induced ROS (Nazio *et al.*, 2019). CSCs can also present an increased **aldehyde dehydrogenase (ALDH) activity**, which participates in the drug detoxification through the oxidation of aldehydes to carboxylic acids, leading to the reduction of ROS levels and thus preventing ROS-induced DNA damage (Raha *et al.*, 2014; Zanoni *et al.*, 2022). To decrease the **oxidative stress** due to intracellular ROS levels, CSCs can increase the expression of antioxidant machinery components, such as the superoxide dismutase or the glutathione reductase, in the different subcellular compartments to scavenge ROS (Das and Roychoudhury, 2014). They generally possess enhanced DNA repair mechanisms to counterbalance the damage inflicted to DNA by ROS (Schulz *et al.*, 2019; Skvortsov *et al.*, 2015). Moreover, some CSCs have the capacity to switch their **metabolism**, between glycolysis and oxidative phosphorylation (OXPHOS), which participates in the modulation of the oxidative stress. As the mitochondrial oxidative phosphorylation is an important source of ROS, the up-regulation of glycolysis coupled with OXPHOS decrease is called the **Warburg effect** and helps maintain the redox (reduction-oxidation) balance upon treatments (Daniel *et al.*, 2021; Movahed *et al.*, 2019). In addition, CSCs can be resistant to oxidative stress by being located within a **hypoxic niche** triggering high levels of HIF factors and enhanced ROS detoxification mechanisms (Kabakov and Yakimova, 2021). Furthermore, CSCs can have increased **survival** mechanisms, such as an efficient **anti-apoptotic** system, an high **telomerase activity** and enhanced **DNA damage repair** mechanisms, such as the base excision repair (BER) (Liu *et al.*, 2006a; Makki *et al.*, 2015; Skvortsov *et al.*, 2015; Wesbuer *et al.*, 2010). Finally, as CSCs in a **quiescent state** have a reduced DNA replication velocity, and are therefore less sensitive to replicative stress, induced by radiations or chemotherapy drugs, creating DNA damage (Carruthers *et al.*, 2018; Skvortsova *et al.*, 2015).

1.5 THERAPEUTIC TARGETING OF CANCER STEM CELLS

1.5.1 CSC MARKERS

The first step to study CSCs and develop therapeutics against them is to properly discriminate them from the rest of the cancer cell population. To do so, specific markers have been found expressed depending on the cancer model and cell lines. Additionally to the expression of the pluripotency factors (e.g. OCT4, SOX2, NANOG, KLF4...), cell surface proteins, named cluster of differentiation (CD), are widely used to identify CSCs. For example, the CD133⁺ cells and CD44⁺/CD24⁻

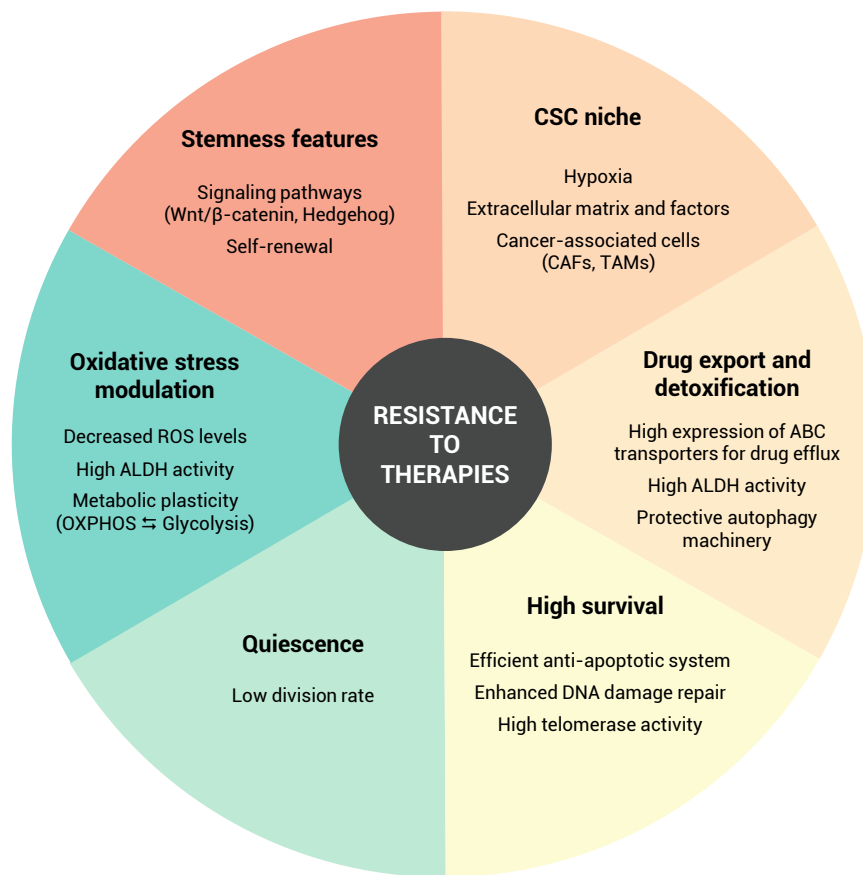


FIGURE 9 RESISTANCE MECHANISMS OF CANCER STEM CELLS. Numerous interdependent mechanisms and actors are involved in the increased resistance of CSCs, such as stemness-associated signaling pathways, hypoxic niche, drug efflux, increased survivability, oxidative stress control, or being in a quiescent state. ABC= ATP-binding cassette; ALDH= aldehyde dehydrogenase; CAFs= cancer-associated fibroblasts; CSC= cancer stem cell; OXPHOS= oxidative phosphorylation; ROS= reactive oxygen species; TAMs= tumor-associated macrophages.

cells have been characterized as the CSC population in glioblastoma and breast cancer respectively. Other CSC-specific mechanisms such as high ALDH activity (ALDH⁺), high drug efflux (Hoechst side population), or low proteasome activity (C-terminal degron of murine ornithine decarboxylase (cODC) fusion protein) are also used in different cancer models, including breast cancer. As the use of one marker is not sufficient to identify the CSC population, they need to be combined and confirmed with functional tests such as sphere-forming capacity (SFC) tests to reveal their self-renewal capacities or in vivo tumorigenicity tests to evaluate their tumorigenic potential. Detailed information about CSC markers used in breast cancer are given in section 3.3.1 “Markers and isolation” at page 101.

1.5.2 STRATEGY

Innovative strategies have been designed, combining both conventional therapies with novel therapies that do not aim to shrink the tumor but to eliminate the CSCs which are known to sustain the long-term tumor growth (Figure 10A and B) (Batlle and Clevers, 2017; Eun *et al.*, 2017).

Thereby, new therapeutic agents are developed to radio-sensitize or chemo-sensitize tumors, by targeting stemness regulators and other resistance mechanisms specific to CSCs, CSC-specific signaling pathways, or the CSC microenvironment.

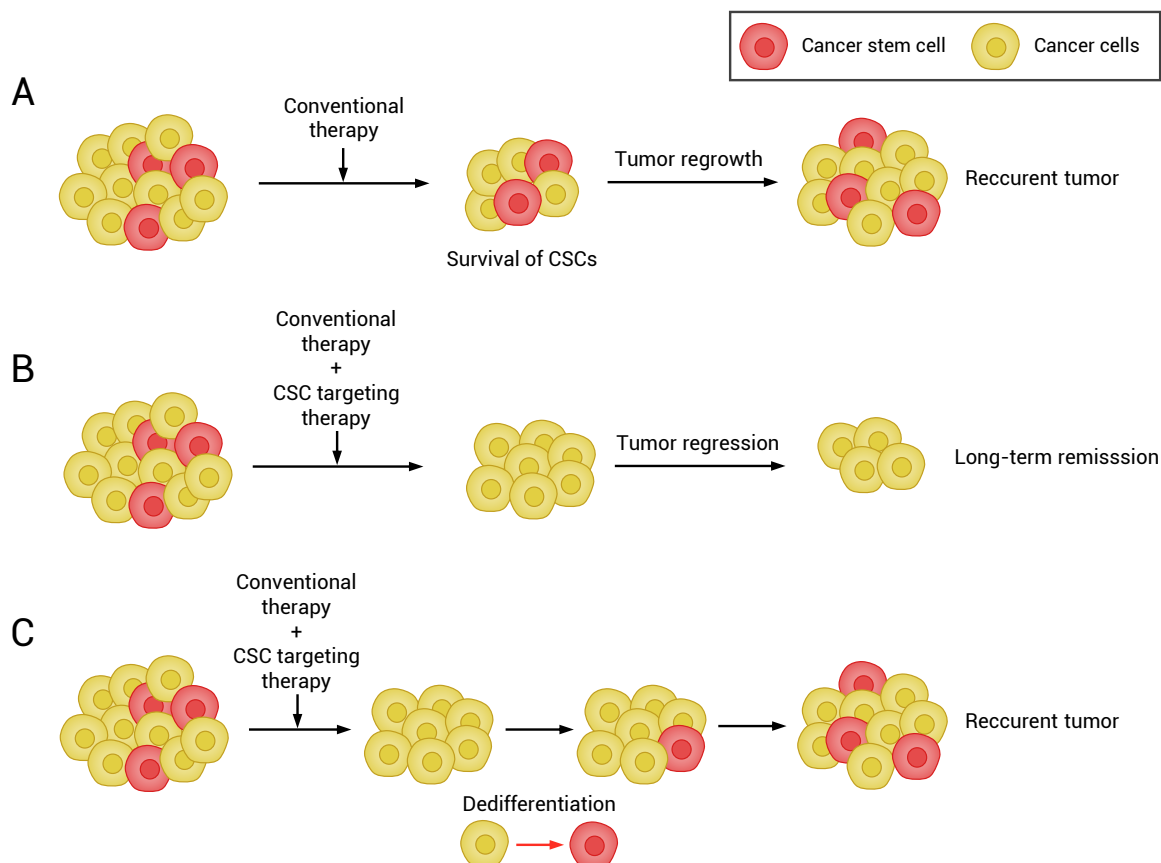


FIGURE 10 THERAPEUTIC STRATEGY MODELS. **A.** The conventional therapies are unable to eliminate CSCs, which cause tumor regrowth and cancer relapse. **B.** A combined therapy, including a conventional treatment and a targeted treatment against CSCs, is used to deplete specifically this population, leading to tumor shrinkage and increasing the probability of a long-term remission. **C.** The major limitation of the previous model is the plasticity of cancer cells and notably their capacity to dedifferentiate back into a CSC state.

1.5.3 APPROACHES

Many approaches to target CSCs have been considered. Among them, first, chemical- or antibody-dependent inhibition of the Wnt/ β -catenin, Notch, and Hedgehog signaling pathways is the most direct one (Clara *et al.*, 2020; Yang *et al.*, 2020b).

For instance, the Notch pathways can be inhibited using γ -secretase inhibitors (GSIs). These inhibitors are the oldest and largest class of agents that target Notch signaling by blocking the cleavage of the Notch intracellular domain (NICD). However, they generally cause off-tumor effects on the gastrointestinal system, as they impair the proliferation of crypt progenitors and induce their differentiation, and affect the generation of lymphocytes, yet they remain effective (Andersson and Lendahl,

2014; Es *et al.*, 2005). One of these inhibitors, the PF-03084014, is being tested in several clinical trials, notably one on advanced-stage triple-negative breast cancer (TNBC) in combination with the docetaxel chemotherapeutic agent (NCT01876251, phase I) (Locatelli *et al.*, 2016).

As another example, smoothened (SMO) antagonists can be used as treatments to inhibit the Hedgehog pathway. For instance, the glasdegib is actually in test on acute myeloid leukemia (AML) (NCT01546038, Phase II and NCT03416179, Phase III) (Cortes *et al.*, 2019; Cortes *et al.*, 2018).

A multitude of other inhibitors have also been tested: inhibitors of epigenetic modifying drugs, such as histone deacetylase (HDAC) inhibitors or DNMT inhibitors (Toh *et al.*, 2017), inhibitors of ABC transporters in combination with chemotherapeutic agents (Hou *et al.*, 2013), inhibitors of the ALDH activity in combination with chemotherapeutic agents or radiations (Kim *et al.*, 2013a), or even inhibitors of the DNA damage response in combination with radiotherapy, such as inhibitors of checkpoint kinases 1 and 2 (Chk1 and Chk2) (Bao *et al.*, 2006).

Another approach, called differentiation therapy, aims to induce the differentiation of CSCs to be able to eliminate them in combination with conventional therapies. For example, the retinoic acid, a derivative of vitamin A, binds to its receptor the retinoic acid receptor (RAR) which triggers its transcriptional activity on differentiation genes. Its pharmaceutical form named all-trans retinoic acid (ATRA) is being tested in several clinical trials on several cancer models (Jin *et al.*, 2017; Thé, 2018)

Furthermore, immunotherapy approaches have also been developed, including the use of antibody-drug conjugates (ADCs), bispecific antibodies, and chimeric antigen receptor T (CAR-T) cells. First, the anti-CSC ADCs are specific antibodies targeting a CSC surface marker conjugated with a cytotoxic drug, with an anti-mitotic activity or DNA damage-inducing capacity to specifically eliminate CSCs (Chalouni and Doll, 2018; Marcucci *et al.*, 2019). Then, immunotherapy using bispecific T cell engager (BiTE) antibodies, antibodies recognizing two antigens, aims to facilitate the action of immune cells against CSCs. Indeed, one antigen is a CSC marker (e.g. anti-CD133), to target CSCs, and the other (e.g. anti-CD3) serves to recruit immune cells such as T cells (Dai *et al.*, 2021; Huang *et al.*, 2013a). Lastly, CAR-T cells are T cells transfected with chimeric antigen receptor targeting CSC-specific antigens (e.g. CD133, ALDH, CD44) to induce a T cell action directed against CSCs (Alhabbab, 2020; Masoumi *et al.*, 2021).

1.5.4 LIMITATIONS

Although these therapeutic approaches relying on CSCs depletion are promising, they come with some limitations.

First, the tumor heterogeneity is applied to the CSC population as well. CSC markers are not expressed at the same time by all CSCs, and the use of a single CSC marker is not sufficient to properly segregate all CSCs from tumor cells (Eun *et al.*, 2017; Huang *et al.*, 2013b; Zheng *et al.*, 2018).

Second, the CSC location in niches within the tumor is important for their sensitivity to drugs. Different CSC states, as proliferative or quiescent, reside in different niches (perivascular or hypoxic niches), making them differently exposed and thereby differently sensitive to the same drug (Marcucci and Corti, 2012; Plaks *et al.*, 2015).

Third, concerning the targeting of signaling pathways activated in CSCs, the abrogation of only one of them could be insufficient as cross-talks and compensatory mechanisms have been described (Jaeger *et al.*, 2017; Sun *et al.*, 2016).

Fourth, as CSCs and normal stem cells share the same stemness-related overexpressed genes and overactivated pathways, in this case the CSC-specific therapies may also induce significant toxicity in normal tissue stem cells. The solution would be to find CSC-specific therapeutic solutions that are not critical for their non-tumorigenic counterparts to lower the risk of side effects (Diehn *et al.*, 2009).

Last, as explained in section 1.2 “Tumor development models” at page 46, tumor cells are plastic and non-CSCs can dedifferentiate into CSCs. Therefore, the depletion of CSCs within the tumor might not be enough for a long-term tumor regression as new CSCs can arise from differentiated cancer cells (Figure 10C) (Das *et al.*, 2020; Gupta *et al.*, 2019; Sanaa, 2022). Moreover, the phenotypic plasticity and capacity of tumor cells to interconvert between a differentiated state and a stem-like state represent a substantial difference compared to normal stem cells, which are almost organized in a unidirectional model where differentiated cells generally cannot convert back to a stem-like state, except when forcing the expression of pluripotent markers to process *in vitro* reprogramming into iPS cells (Hanahan, 2022). **Accordingly, targeting the cancer plasticity, especially dedifferentiation mechanisms, might be the key to overcoming the cancer resistance to therapies and reducing the risk of cancer recurrence.**

KEY POINTS

- Numerous resistance mechanisms can be upregulated in CSCs, such as drug efflux, oxidative stress modulation, protective autophagy, anti-apoptotic system, DNA damage repair, or metabolic plasticity (Figure 9).
- As CSCs sustain the tumor's long-term growth, therapeutic strategies have been developed to eliminate CSCs and sensitize the tumor to therapies, by combining a conventional therapy with a CSC-specific therapy (Figure 10B).
- The CSC-specific therapies can target CSC-related markers, features, resistance mechanisms, or microenvironment using several approaches such as signaling pathways inhibitors, epigenetic modifying enzymes inhibitors, drug efflux pumps inhibitors, treatment inducing CSC differentiation, or immunotherapy approaches (e.g. ADC, BiTE, CAR-T cells).
- Although these approaches are promising, the main limitation to the CSC-specific targeting is their plasticity. Indeed, new CSC can arise from the dedifferentiation of bulk cancer cells, which makes the CSCs complete depletion impossible and cause cancer recurrence (Figure 10C). In addition, as one major difference compared to normal cells is the plasticity of cancer cells and their capacity to interconvert between CSC and non-CSC states, the dedifferentiation mechanisms involved should be cancer-specific. Accordingly, targeting the dedifferentiation mechanisms might provide an even more effective therapeutic solution to overcome cancer resistance.

[^ Back to Table of Contents](#)

2

PLASTICITY AND DEDIFFERENTIATION INTO CANCER STEM CELLS

☰ CHAPTER CONTENTS

2.1 PHENOTYPIC PLASTICITY	65
2.1.1 DEFINITIONS	65
2.1.2 CSC PLASTICITY AND EMT	68
📌 KEY POINTS	70
2.2 INDUCTION OF DEDIFFERENTIATION	70
2.2.1 SPONTANEOUS DEDIFFERENTIATION	72
2.2.2 THERAPY-INDUCED DEDIFFERENTIATION	73
2.2.3 FORCED REEXPRESSION OF PLURIPOTENCY FACTORS	75
📌 KEY POINTS	76
2.3 REGULATION OF DEDIFFERENTIATION	76
2.3.1 HYPOXIA	76
2.3.2 EXTRACELLULAR VESICLES	78
2.3.3 INFLAMMATION AND SECRETED MOLECULES	79
2.3.4 EPIGENETICS	81
📌 KEY POINTS	85

⬆️ [Back to Table of Contents](#)

2.1 PHENOTYPIC PLASTICITY

2.1.1 DEFINITIONS

The **plasticity** is the ability for a cell to reversibly assume different cellular phenotypes also referred to as identities or states. These changes of phenotypes are carried out throughout **reprogramming events**. Consequently, the **CSC plasticity** is the ability to switch between the CSC and non-CSC states.

Three types of reprogramming events can occur within the CSC plasticity: ① **differentiation**, from a pluripotent state towards a differentiated one, ② **dedifferentiation**, from a differentiated state towards a pluripotent one, also called **reprogramming**, and ③ **transdifferentiation**, from a differentiated state in one lineage to another lineage, also called **direct reprogramming** (Figure 11) (Eguizabal *et al.*, 2013; Hanahan, 2022; Yamada *et al.*, 2014).

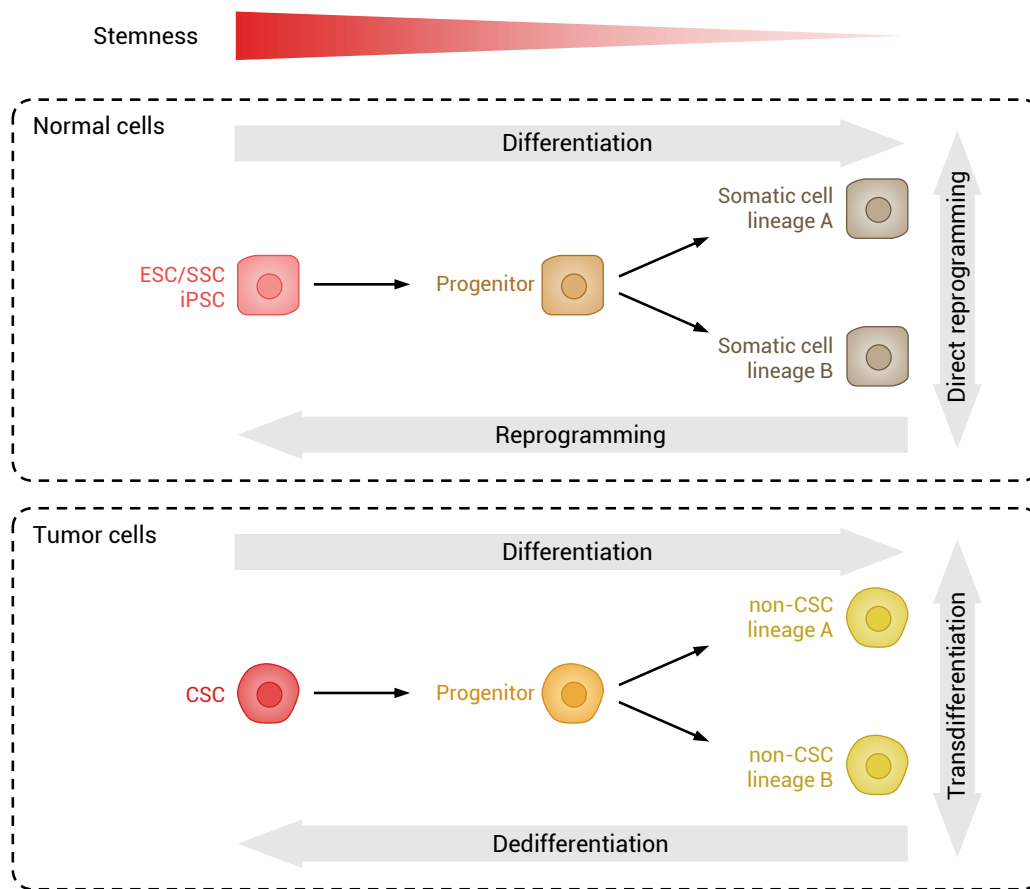


FIGURE 11 PLASTICITY AND REPROGRAMMING EVENTS. **CSC**= cancer stem cell; **ESC**= embryonic stem cell; **iPSC**= induced pluripotent stem cell; **non-CSC**= cancer non-stem cell; **SSC**= somatic stem cell. (Granados *et al.*, 2020; Hanahan, 2022; Yamada *et al.*, 2014)

In normal cells, the **reprogramming** term is used to designate the resetting of epigenetic marks during early embryogenesis and germ cell formation allowing the reexpression of gene programs associated with pluripotency, enabling the cell to acquire a stem cell phenotype (Xavier *et al.*, 2019; Zeng and Chen, 2019).

The **artificial reprogramming** to a pluripotent state was first discovered *in vitro* by forcing the reexpression of the four transcription factors OCT4, SOX2, C-MYC, and KLF4 (Takahashi *et al.*, 2007; Takahashi and Yamanaka, 2006). In this context, the reprogrammed cells with stemness properties are called induced pluripotent stem (iPS) cells.

Additionally, the **direct reprogramming** term has been used to describe the cell phenotype conversion from one lineage into another, without undergoing an intermediate pluripotent state, also called transdifferentiation (Wang *et al.*, 2021).

The reprogramming toward pluripotency, or dedifferentiation, is an unusual event in normal cells, which have been described mainly for tissue repair purposes, as they are unidirectionally organized toward terminal differentiation. In cancer, cells are more plastic as they are able to dediffer-

entiate back to a stem state with pluripotency and self-renewal features enabling tumor growth. This unlocking of phenotypic plasticity has been added to the hallmarks of cancer in 2022 by Hanahan, as described in Figure 12, completing the established ones from 2000 and 2011 (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). Hanahan (2022) describes four reprogramming types of events within this phenotypic plasticity: differentiation, dedifferentiation, blocked differentiation, and transdifferentiation.

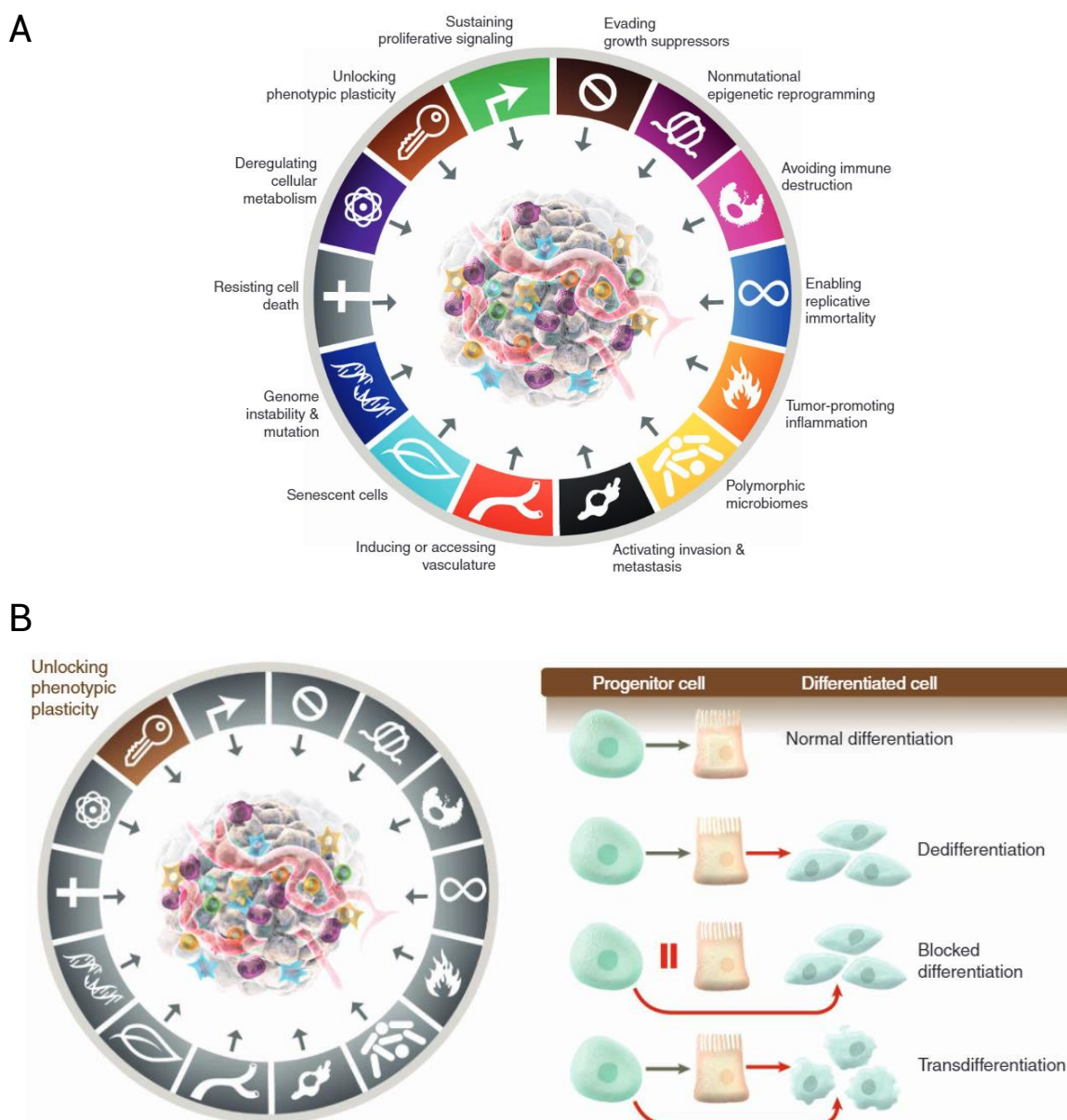


FIGURE 12 NEW ADDITION TO HALLMARKS OF CANCER: UNLOCKING PHENOTYPIC PLASTICITY. **A.** Hallmarks of cancer. **B.** Unlocking phenotypic plasticity hallmark. Figure from Hanahan (2022)

In cancer cells, the **dedifferentiation**, also called **reprogramming**, **retrodifferentiation**, or **non-CSC-to-CSC conversion**, is the reverse process of differentiation, in which the differentiated cells (non-CSCs) with specialized functions become more undifferentiated and gain the self-renewal

ability and a pluripotent state (CSCs) (Cabillic and Corlu, 2016; Marjanovic *et al.*, 2013).

The **transdifferentiation** have been described in cancer cells as they can dynamically interconvert their phenotypes and exhibit different lineage characteristics, without going through a pluripotent state in between (Yuan *et al.*, 2019).

The **blocked differentiation** describes the incompletely differentiated progenitors, that cannot go further in the differentiation to maintain a high proliferation (Hanahan, 2022).

2.1.2 CSC PLASTICITY AND EMT

The epithelial-to-mesenchymal transition (EMT) is defined as a reversible phenotypic switch from an epithelial state to a mesenchymal state. The cancer cell loses its epithelial characteristics, such as cell junctions and apical-basal polarity, and gains mesenchymal characteristics such as an elongated morphology, migration, and invasion capacities. Thus, the cell can migrate through surrounding tissues to blood vessels, disseminate and colonize a secondary site by forming a new tumor, called metastasis, while undergoing the reverse process, the mesenchymal-to-epithelial transition (MET) (Kalluri and Weinberg, 2009; Yilmaz and Christofori, 2009).

The EMT is mediated by a set of EMT-inducing transcription factors (EMT-TFs) which are divided into three main families of proteins, the Snail family transcriptional repressor (SNAIL) family (e.g. SNAIL1 and SNAIL2, also known as Snail and Slug), the zinc finger E-box binding homeobox (ZEB) family (e.g. ZEB1 and ZEB2) and the Twist family BHLH transcription factor (TWIST) family (e.g. TWIST1 and TWIST2) (Craene and Berx, 2013).

As both the EMT process and the CSC state involve phenotypic plasticity and are held responsible for the metastatic process, they have been intricately linked (Polyak and Weinberg, 2009; Scheel and Weinberg, 2012). Several studies have observed that cells undergoing EMT acquire stemness properties (Mani *et al.*, 2008; Masui *et al.*, 2014; Morel *et al.*, 2008; Rhim *et al.*, 2012; Shuang *et al.*, 2014; Talati *et al.*, 2015; Xie *et al.*, 2011). For instance, in breast cancer cells, the induction of EMT, by EMT-TFs or cytokines, led to increased tumorigenic properties and expression of CSC markers (Mani *et al.*, 2008; Morel *et al.*, 2008; Xie *et al.*, 2011). It has been described that ZEB1 promotes the EMT as well as SOX2 and KLF4 pluripotency factors, and is required for stemness and tumorigenic capacities (Krebs *et al.*, 2017; Wellner *et al.*, 2009). Hence, the activation of the EMT program is associated with the acquisition of stem-like characteristics.

Although publications often depict the EMT program as the dedifferentiation of cancer cells, where non-CSCs are associated with epithelial features (epithelial/non-CSC state) and CSCs with mesenchymal features (mesenchymal/CSC state), the link between CSC plasticity and EMT is more complex (Junk *et al.*, 2013; Smigiel *et al.*, 2017; Smigiel *et al.*, 2018). In fact, during the EMT process, in between the initial epithelial state and the terminal mesenchymal state, the cell undergoes a variety of intermediate steps corresponding to hybrid epithelial/mesenchymal (E/M) phenotypes

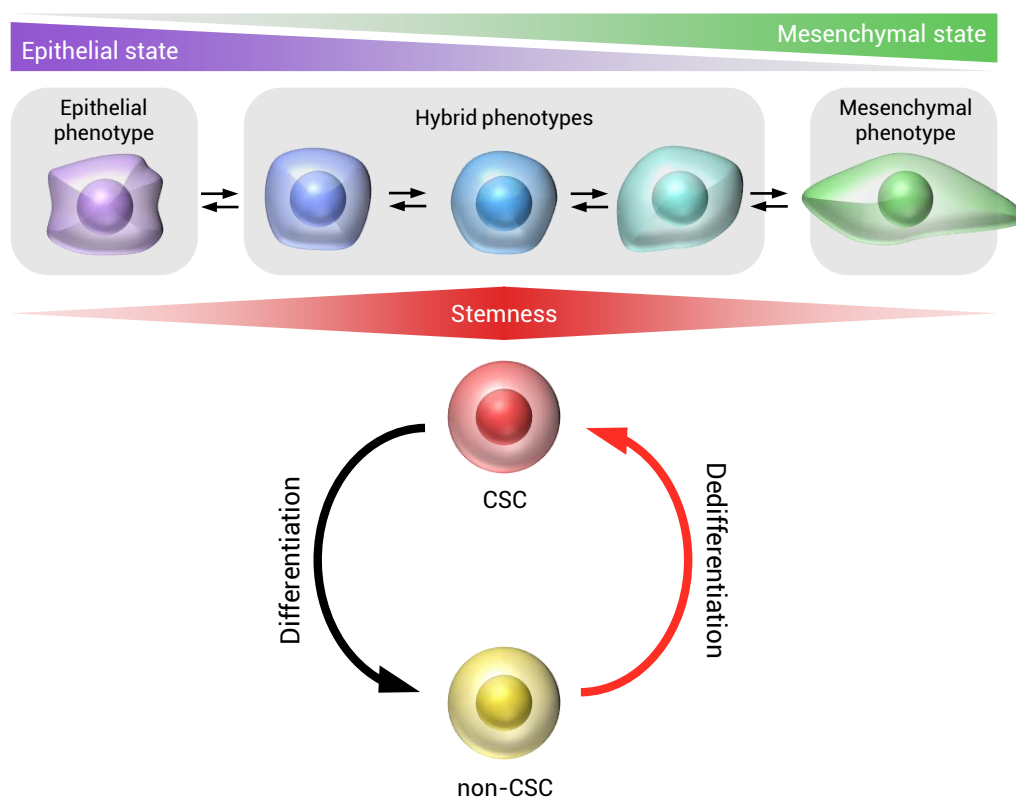


FIGURE 13 CANCER CELL PLASTICITY: RELATION BETWEEN CSC PLASTICITY AND EMT. (Fanelli *et al.*, 2020; Silva-Diz *et al.*, 2018)

(Figure 13) (Gupta *et al.*, 2019; Jolly *et al.*, 2018). These hybrid cells that have undergone a partial EMT exhibit stem cell properties, and are thereby less differentiated than those on either end of the EMT spectrum (Bierie *et al.*, 2017; Fanelli *et al.*, 2020; Silva-Diz *et al.*, 2018; Thankamony *et al.*, 2020). Pastushenko *et al.* (2018) demonstrate the presence of multiple EMT stages that constitute different tumor subpopulations, with different transcriptional and epigenetic signatures, and different invasiveness and metastatic capacities. Kröger *et al.* (2019) found that the tumorigenic capacity of breast cancer cells was maintained by the hybrid E/M phenotype of the CD104⁺/CD44^{high} cells, and lost when they become highly mesenchymal. They identified the canonical Wnt signaling pathway as a key mediator of this hybrid state, activated by the EMT-TFs, but replaced by the non-canonical pathway when the cell acquires the ZEB1-driven mesenchymal phenotype (Kröger *et al.*, 2019). Consequently, multiple CSC states can exist within the EMT spectrum, characterized by epithelial and mesenchymal features. Liu *et al.* (2013c) found epithelial-like breast CSCs, proliferative CSCs with a high ALDH activity, and mesenchymal-like breast CSCs, relatively quiescent CSCs mediating tumor invasion in blood vessels, and suggest that the CSCs can transit back-and-forth between those states depending on tumor microenvironment signals (Liu *et al.*, 2013c).

To conclude, the EMT program participates in the generation of CSCs, mediated by the tumor microenvironment, especially upon treatment, thus it contributes to the CSC pool expansion and

resistance to therapies. It is not clearly stated in the literature whether or not the EMT should be considered as a dedifferentiation or transdifferentiation process. Furthermore, the acquisition and loss of differentiated features during the EMT involve CSC plasticity mechanisms allowing the cell to access different cellular identities by going through a hybrid E/M plastic intermediate CSC state in between. Above all, this interconnection between the epithelial-mesenchymal plasticity and the CSC plasticity supports the very dynamic and transitory change of states of cancer cells, which drives the cancer development, metastasis formation, and resistance to therapies.

KEY POINTS

- The cancer stem cell (CSC) plasticity is the ability for a cell to switch between the CSC and non-CSC phenotypes. Different phenotypic reprogramming events can occur: the differentiation, the dedifferentiation (or reprogramming), and the transdifferentiation (or direct reprogramming).
- The epithelial-to-mesenchymal transition (EMT) also involves a phenotypic plasticity as it is characterized by a conversion from an epithelial phenotype to a mesenchymal one. During the EMT process, the cell undergoes a variety of intermediate steps, referred to as hybrid E/M phenotypes.
- Interestingly, hybrid E/M cells are less differentiated than those on either end of the EMT spectrum, as they exhibit stem cell properties. Combined with the CSC plasticity, it indicates that multiple CSC states can exist within the EMT spectrum, and that the EMT process is intricately linked with the emergence of cells presenting a CSC phenotype.

[^ Back to Table of Contents](#)

2.2 INDUCTION OF DEDIFFERENTIATION

The dedifferentiation of non-CSC into CSC is a phenotypic switch from a differentiated state to a stem-like state of cancer cells. This process leads to the *de novo* generation of CSCs, participating in the CSC pool maintenance or expansion. It was demonstrated that the dedifferentiation events occur spontaneously within the tumor, but can also be induced by several extrinsic or intrinsic factors, especially following anti-cancerous treatments, in hypoxic areas of the tumor, or upon stimuli from other cells within the tumor. The [Table 1](#) lists studies reporting for the conversion of non-CSC into CSC in different models, with different inducers.

To distinguish the dedifferentiation events from self-renewal of CSCs through cell division of pre-existing ones, the key in dedifferentiation experiments is either to tag the CSC and non-CSC to track phenotypic switches, or to deplete the population in CSCs beforehand, so the observed proportion of CSCs can only come from the reprogramming of non-CSCs. Yet, one limit of this experimental strategy resides in the difficulty to fully discriminate the CSC population as no CSC marker alone is sufficient. In the case of fluorescence-activated cell sorting (FACS) sorting, the purified non-CSC population might contain residual CSCs. However, the self-renewal of a few CSCs wouldn't be sufficient to explain the increased proportion of CSCs.

TABLE 1 - First page

INDUCER	TYPE OF CANCER	EXPERIMENTAL MODEL	CSC MARKERS USED	REGULATORY MECHANISMS UNCOVERED	REFERENCE
Spontaneous	Breast cancer	Transformed human mammary epithelial cells	CD44 ⁺		Chaffer <i>et al.</i>, 2011
Spontaneous	Colon cancer Breast cancer	SW620 cell line MCF-7 cell line	CD133 ⁺ CD44 ⁺ CD24 ⁻		Yang <i>et al.</i>, 2012
Chemotherapy (carboplatin)	Hepatocellular carcinoma	HepG2 and Huh7 cell lines	Hoechst side population		Hu <i>et al.</i>, 2012
Chemotherapy (temozolomide)	Glioblastoma	Cell lines : U251, U87, A172 PDX lines : GBM43, GBM6, GBM26	CD133 ⁺ CD15 ⁺	Hypoxia involvement	Auffinger <i>et al.</i>, 2014
Chemotherapy (temozolomide)	Glioblastoma	Cell lines : U251 and U87 PDX lines : GBM43 and GBM6	CD133 ⁺ SOX2-p-RFP ⁺ OCT4-p-RFP ⁺ NANOG-p-RFP ⁺	HIF-1 α and HIF-2 α involvement	Lee <i>et al.</i>, 2016a
Chemotherapy (5-fluorouracil, doxorubicin, cyclophosphamide)	Breast cancer	Patient tumors and MCF-7 cell line	CD44 ⁺ CD24 ⁻	NF- κ B-dependent IL6 inflammatory feedback loop activation	Saha <i>et al.</i>, 2016
Radiotherapy (2, 4 Gy)	Hepatocellular carcinoma	HepG2 and Huh7 cell lines	Hoechst side population	SOX2 and OCT4 requirement	Ghisolfi <i>et al.</i>, 2012
Radiotherapy (4, 8 Gy)	Breast cancer	SUM159PT, MCF-7 and T47D cell lines	ALDH ⁺ CD44 ⁺ CD24 ⁻ and low proteasome activity	Re-expression of OCT4, SOX2, NANOG, and KLF4, Notch signaling activation	Lagadec <i>et al.</i>, 2012
Radiotherapy (fractionated 3.75 Gy/day \times 5 days)	Breast cancer	MDA-MB-231, SUM149, UACC-812 and 4T1 (mouse) cell lines	ALDH ⁺	NF- κ B signaling activation	Wang <i>et al.</i>, 2014a
Radiotherapy (8 Gy)	Breast cancer	SUM159PT and MDA-MB-231 cell lines	ALDH ⁺	IL6-JAK2-STAT3 signaling activation	Arnold <i>et al.</i>, 2020b
Radiotherapy (8 Gy)	Glioblastoma	Primary human glioblastoma cell lines and PDX	Low proteasome activity	Akt signaling activation	Bhat <i>et al.</i>, 2020
Radiotherapy (2, 4, 8, 10, and 20 Gy)	Pancreatic cancer	PaTu8988 cell line, tumor specimens and PDX	CD133 ⁺	HMGB1-TLR2 interaction regulating the YAP/HIF-1 α signaling	Zhang <i>et al.</i>, 2019
HDAC inhibitors (Valproic acid)	Breast cancer	SUM159PT and MDA-MB-231 cell lines	ALDH ⁺	Wnt/ β -catenin signaling activation	Debeb <i>et al.</i>, 2012

TABLE 1 - Second page

INDUCER	TYPE OF CANCER	EXPERIMENTAL MODEL	CSC MARKERS USED	REGULATORY MECHANISMS UNCOVERED	REFERENCE
Hypoxia	Glioblastoma	T4121 cell line, brain tumor patient specimens and human glioblastoma xenografts	CD133 ⁺	HIF-2 α involvement	Heddleston et al., 2009
Hypoxia	Pancreatic cancer	Panc-1 and BxPC-3 cell lines	CD133 ⁺	HIF-1 α involvement	Zhu et al., 2013
Hypoxia	Glioblastoma Hepatocellular carcinoma Lung cancer	GL261 cell line HepG2 cell line A549 cell line	CD133 ⁺ CD15 ⁺ NESTIN ⁺		Wang et al., 2017d
Hypoxia	Glioblastoma	GL261 and U87 cell lines	CD133 ⁺ CD15 ⁺ NESTIN ⁺	HIF-1 α involvement	Wang et al., 2017c
Fibroblast-derived extracellular vesicles coupled with chemotherapy (5-fluorouracil or oxaliplatin)	Colorectal cancer	HT-29 and SW620 cell lines	CD133 ⁺ ALDH ⁺	Fibroblast-secreted exosomal Wnts activate Wnt/ β -catenin signaling	Hu et al., 2019b
CSC-derived exosomes	Glioblastoma	WJ1, WJ2, U251 and U87 cell lines	WJ2 glioblastoma CSC cell line		Sun et al., 2020

TABLE 1 MAIN STUDIES REPORTING THE DEDIFFERENTIATION OF NON-CSCs INTO CSCs. ALDH= aldehyde dehydrogenase; CAF= cancer-associated fibroblast; CD= cluster of differentiation; Gy= Gray; HDAC= histone deacetylase; HIF= hypoxia-inducible factor; HMGB1= high-mobility group box 1; IL6= interleukin 6; JAK2= Janus-activated kinase 2; OCT4= octamer-binding transcription factor 4; PDX= patient-derived xenograft; SOX2= sex determining region Y-box 2, STAT3= signal transducer and activator of transcription 3; TLR2= Toll-like receptor 2; YAP= Yes1 associated transcriptional regulator.

2.2.1 SPONTANEOUS DEDIFFERENTIATION

The conversion between non-stem state and stem state is a process occurring spontaneously within the tumor. Indeed, [Chaffer et al. \(2011\)](#) demonstrated by FACS isolation of transformed cells that both the non-stem fraction and the CSC fraction of cells can give rise to one another *in vitro*. The conversion of non-CSCs into CSCs also occurs *in vivo*, as the generated tumors after mice injection presented a significant proportion (16%) of CSCs after 8-10 weeks, and it also highlights the importance of tumor microenvironment in the arising of CSC. Another study on colon cancer cells and breast cancer cells revealed the acquisition of stem cell markers and the appearance of new CSCs, in the CSC depleted population ([Yang et al., 2012](#)). The non-CSC fraction and the CSC fraction were purified and, after separated cultures of both populations, they exposed an inter-conversion equilibrium whereby both compartments can convert to one another until an intrinsic balance is found

with a CSC proportion stabilized around the CSC proportion of the unsorted population.

Several studies have shown that, after cell sorting or tagging of non-CSC population, they can convert into CSC without any inducer (Auffinger *et al.*, 2014; Debeb *et al.*, 2012). In a mixed population, the GFP-tagged non-CSCs give rise to around 1 to 2% of CSCs (CD133⁺ CD15⁺ cells) without induction in the glioblastoma model, demonstrating a basal dedifferentiation process occurring spontaneously to restore an equilibrium proportion of CSC (Auffinger *et al.*, 2014).

Moreover, the dedifferentiation events and the state equilibrium between the CSC and non-CSC subpopulation have been confirmed by mathematical models of cancer kinetics to apprehend the population dynamics (Butner *et al.*, 2022; Gupta *et al.*, 2011; Jilkine, 2019; Zapperi and Porta, 2012).

2.2.2 THERAPY-INDUCED DEDIFFERENTIATION

The resistance to therapy has been associated with an increase of the CSC proportion in several tumor models (Abubaker *et al.*, 2013; Lagadec *et al.*, 2010; Lee *et al.*, 2011; Lu *et al.*, 2015; Murata *et al.*, 2019; Wang *et al.*, 2013a). To explain the CSC expansion upon treatments, Auffinger *et al.* (2014) postulated three scenarios co-occurring upon treatment: ① selection of CSCs over non-CSCs due to their high resistance to therapies, ② expansion of the CSC pool caused by a shift from asymmetric division towards symmetric division, and finally ③ dedifferentiation of non-CSCs, resulting in the generation of non-pre-existing CSCs. The two first scenarios are describing a CSCs enrichment from pre-existing CSC, while in the third scenario CSC arises *de novo*, and their co-occurrence upon treatment is confirmed by kinetic modelizations (Gao *et al.*, 2014). Hence, the treatment-induced dedifferentiation and underlying mechanisms have been particularly investigated (Chen *et al.*, 2017c; Li *et al.*, 2016).

CHEMOTHERAPY-INDUCED DEDIFFERENTIATION

In **hepatocellular carcinoma**, Hu and Fu (2012) treated the sorted non-CSC population with carboplatin, and maintained them in stem cell media and ultra-low attachment plates. The treated non-CSCs showed an increase in self-renewal capacities, sphere formation, and expression of SOX2 and OCT4, compared to untreated ones.

In **glioblastoma**, Auffinger *et al.* (2014) investigated the formation of new CSCs from non-CSCs upon temozolomide treatment. In the sorted and GFP-tagged non-CSC pool of cells, around 5% of them expressed stemness markers 8 days after temozolomide treatment, which is significantly higher than in the untreated control. *In vivo*, the non-CSC population previously treated with temozolomide was able to form tumors in mice which presented an increase of cells expressing CSC markers compared to the untreated control. Interestingly, after temozolomide exposure, they found increased levels of HIF-1 α and HIF-2 α in the newly formed CSCs, and more intratumoral hypoxic areas in xenografts, matching the hypoxia and HIFs mediated maintenance of stemness, but also

highlighting their regulatory role in the temozolomide-induced dedifferentiation of non-CSC into CSC in the glioblastoma. This role was further confirmed by [Lee et al. \(2016a\)](#), through real-time imaging of glioblastoma cell lines and PDX lines treated with temozolomide, using glioblastoma CSC-specific promoter-based reporter systems. They reported an increased rate of phenotypic switches from non-CSCs to CSCs at the single-cell level after temozolomide exposure, alongside an increased expression of both HIF-1 α and HIF-2 α . As the knockdown of HIFs expression reduced the temozolomide-induced non-CSC-to-CSC conversion, it indicates that HIFs have a crucial role in the dedifferentiation of glioblastoma cancer cells.

In **breast cancer**, although the purified non-CSC population failed to convert into CSCs upon chemotherapy, in a mixed population of CSCs and non-CSCs, the tagged non-CSCs were able to dedifferentiate into CSCs after chemotherapy ([Saha et al., 2016](#)). Hence, it highlights the requirement for pre-existing CSCs to drive the CSC pool expansion through non-CSC dedifferentiation, after chemotherapy exposure. Indeed, the chemotherapy treatment, composed of 5-fluorouracil, doxorubicin, and cyclophosphamide, triggers a positive inflammatory feedback loop in pre-existing CSCs promoting the non-CSC-to-CSC conversion. This loop relies on the NF- κ B-dependent IL-6 activation ([Iliopoulos et al., 2011](#); [Saha et al., 2016](#)).

RADIOTHERAPY-INDUCED DEDIFFERENTIATION

In **breast cancer**, several studies have demonstrated that CSCs can arise *de novo* from the sorted non-CSC population after irradiation ([Arnold et al., 2020b](#); [Bhat et al., 2020](#); [Bidan et al., 2019](#); [Lagadec et al., 2012](#); [Wang et al., 2014a](#)). The non-CSC population irradiated at 8 Gy has a higher CSC percentage compared to the non-irradiated control 5 days post-irradiation ([Arnold et al., 2020b](#); [Bhat et al., 2020](#); [Bidan et al., 2019](#); [Lagadec et al., 2012](#)). Interestingly, in a mixed population of tagged non-CSCs and CSCs, the generation of CSCs through radio-induced dedifferentiation of non-CSCs is diminished, indicating a control of the pre-existing CSCs over the CSC pool expansion ([Lagadec et al., 2012](#)). Additionally, the inhibition of Notch signaling ([Lagadec et al., 2012](#)), NF- κ B signaling ([Wang et al., 2014a](#)), or STAT3 signaling ([Arnold et al., 2020b](#)), reduced the emergence of new CSCs.

In **pancreatic cancer**, [Zhang et al. \(2019\)](#) found that radio-induced cell death promotes the dedifferentiation of non-CSCs into CSCs. The 7-days co-culture of sorted non-CSCs with lethally irradiated cells (20 Gy) induced a significant increase of CSCs and expression of stemness markers, compared to the co-culture with non-irradiated control cells (0 Gy). Furthermore, they observed that among the damage associated molecular patterns (DAMPs) released by dying cells following radiotherapy, the high-mobility group box 1 (HMGB1) binds to the Toll-like receptor 2 (TLR2) receptor of non-CSCs and activates the Yes1 associated transcriptional regulator (YAP)/HIF-1 α signaling, mediating the dedifferentiation. The inhibition of these regulators individually reduces the dedifferentiation of pancreatic cancer cells and their tumorigenic capacities *in vivo* ([Zhang et al., 2019](#)).

In **hepatocellular carcinoma**, the isolated non-CSC population irradiated at 2 or 4 Gy doses showed increased self-renewal and sphere formation capacities after 14 days post-irradiation of culture in stem cell media, indicating the dedifferentiation of cells (Ghisolfi *et al.*, 2012). And, in **glioblastoma**, the non-CSC population was also able to dedifferentiate after radiotherapy, but not when the Akt signaling was inhibited by a trifluoperazine treatment (Bhat *et al.*, 2020).

HDAC INHIBITORS INDUCED DEDIFFERENTIATION

Histone deacetylase (HDAC) inhibitors, such as valproic acid (VPA), trichostatin A (TSA), and suberoylanilide hydroxamic acid (SAHA) also known as Vorinostat, are used as anticancerous agents as they cause histone hyperacetylation, leading to chromatin remodeling, and thus have growth-inhibitory effects on cancer cells (Wawruszak *et al.*, 2019; Zhang *et al.*, 2017a). In addition, numerous studies used HDAC inhibitors as differentiating agents to specifically target CSCs (Alvarez *et al.*, 2015; Botrugno *et al.*, 2009; Dvorakova and Vanek, 2016; Lin *et al.*, 2018).

However, some studies also demonstrated the opposite effect of HDAC inhibitors, stimulating the dedifferentiation of cancer cells by activating stemness-related pathways. Besides, VPA is known to enhance self-renewal and expansion of hematopoietic stem cells (HSCs) (Bug *et al.*, 2005; Felice *et al.*, 2005; Young *et al.*, 2004).

In breast cancer, VPA was found to selectively radio-sensitize differentiated cells while protecting mammospheres cultures from radiations and promoting their self-renewal (Debeb *et al.*, 2010), while TSA and SAHA induce the expression of stem cell and EMT markers in prostate cancer cells (Kong *et al.*, 2012).

More interestingly, Debab *et al.* (2012) demonstrated that HDAC inhibitors induce the dedifferentiation of breast cancer cells. The sorted non-CSC population treated for 7 days with either VPA, TSA, or SAHA showed increased proportions of CSCs and increased tumorigenic capacities *in vivo* compared to the non-treated population. Moreover, they reported an up-regulation of the Wnt/ β -catenin signaling pathway and its inhibition abolished the HDAC inhibitors-induced over-expansion of CSCs.

2.2.3 FORCED REEXPRESSION OF PLURIPOTENCY FACTORS

Lastly, the induced overexpression of pluripotency factors has been shown to increase the stem-like properties of cells from different cancer types (Chiou *et al.*, 2010; Herreros-Villanueva *et al.*, 2013; Yin *et al.*, 2015). Although no studies describe a dedifferentiation process using populations depleted in CSCs or tagged cells to observe phenotypic conversions, several ones describe populations acquiring stem cell properties induced by pluripotency factors re-expression (Fujiwara *et al.*, 2020; Oshima *et al.*, 2014; Suvà *et al.*, 2014). For instance, the retrovirus-mediated forced expression of OCT4, SOX2, and KLF4 in colon cancer cells or osteosarcoma cells gives rise to a population of cells with a stem-like phenotype (Fujiwara *et al.*, 2020; Oshima *et al.*, 2014). Induced

expression of OCT4, NANOG and SOX2 in glioblastoma cells lead to the dedifferentiation into CSCs, along with the activation of Notch and Wnt/ β -catenin pathways (Olmez *et al.*, 2015). So, in the same way, as somatic cells can be reprogrammed in iPS cells, differentiated cancer cells can also be reprogrammed to acquire a stem-like phenotype through the forced expression of pluripotency factors.

KEY POINTS

- The dedifferentiation of non-CSCs into CSCs can be spontaneous or induced by several extrinsic or intrinsic factors, leading to the resurgence of the CSC population.
- Within the tumor cell population, it exists an inter-conversion equilibrium whereby the CSC and non-CSC compartments can convert to one another, without any inducers, until an intrinsic balance is reached at a stabilized proportion of CSCs.
- The CSC enrichment following therapies is in part caused by the non-CSCs conversion to CSCs, induced by the therapeutic stress, such as chemotherapeutic agents, radiations, and HDAC inhibitors treatments.

^ [Back to Table of Contents](#)

2.3 REGULATION OF DEDIFFERENTIATION

An overview of the reported regulatory mechanisms and mediators participating in the non-CSC-to-CSC conversion is displayed in [Figure 14](#).

2.3.1 HYPOXIA

The hypoxic environment, which is strongly associated with the CSC niche, has also been reported as promoting the CSC plasticity ([Figure 14](#)) (Lee *et al.*, 2016b; Wijaya, 2019).

First, the chemotherapy-induced hypoxia in glioblastoma causes non-CSCs dedifferentiation events as mentioned previously in section 2.2.2 “Chemotherapy-induced dedifferentiation” at page 73 (Auffinger *et al.*, 2014; Lee *et al.*, 2016a). Indeed, the temozolomide treatment induces more intratumoral hypoxic areas and promotes the conversion of non-CSCs to CSCs, with high levels of HIF-1 α and HIF-2 α , a conversion which is impaired when HIFs are inhibited (Auffinger *et al.*, 2014; Lee *et al.*, 2016a).

Then, outside of the chemotherapy exposure context, an hypoxic environment and hypoxia-inducible factors (HIFs) are known to induce an enrichment in CSCs in several models including breast cancer (Helczynska *et al.*, 2003; Louie *et al.*, 2010; Schwab *et al.*, 2012; Semenza, 2016a; Semenza, 2016b; Xie *et al.*, 2016) and glioblastoma (Bar *et al.*, 2010; Bonnin *et al.*, 2017; Hashimoto *et al.*, 2011; Soeda *et al.*, 2009). Indeed, HIFs contributes to CSC induction and maintenance *via* the glucose metabolism regulation to maintain the redox homeostasis under hypoxic conditions, and *via* the signaling pathways activation, including Notch, leading to the expression of pluripotency genes (Gustafsson *et al.*, 2005; Qiang *et al.*, 2011; Semenza, 2016a; Semenza, 2016b).

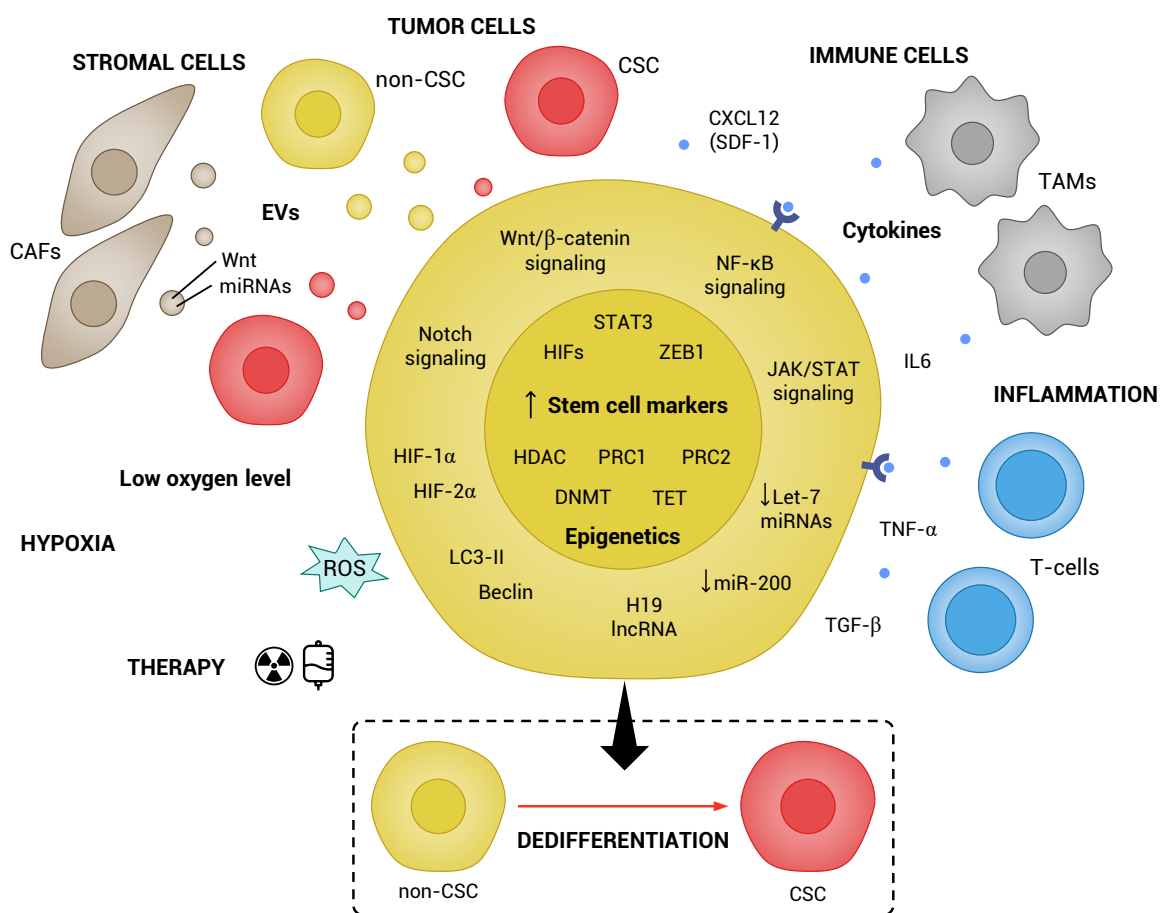


FIGURE 14 REGULATORS AND MECHANISMS INVOLVED IN THE DEDIFFERENTIATION OF NON-CSCS INTO CSCs Regulatory mechanisms and principal mediators involved in the phenotypic switch from the non-CSC state into the CSC state. **ALDH**= aldehyde dehydrogenase; **CAF**= cancer-associated fibroblast; **CSC**= cancer stem cell; **CXCL**= C-X-C motif chemokine ligand; **DNMT**= DNA methyltransferase; **EV**= extracellular vesicles; **HDAC**= histone deacetylase; **HIF**= hypoxia-inducible factor; **IL**= interleukin; **LC3-II**= light chain 3-II; **lncRNA**= long non coding RNA; **miRNA**= microRNA; **non-CSC**= cancer non-stem cell; **PRC**= polycomb repressive complex; **ROS**= reactive oxygen species; **STAT**= signal transducer and activator of transcription; **TAM**= tumor-associated macrophage; **TET**= ten-eleven translocation; **TGF**= tumor growth factor; **TNF**= tumor necrosis factor; **ZEB**= zinc finger E-box binding homeobox.

The hypoxia-mediated CSCs expansion also involves the dedifferentiation of non-CSCs. [Heddleston et al. \(2009\)](#) have demonstrated the promotion of the non-CSC-to-CSC conversion of glioblastoma cells under hypoxic conditions. The sorted non-CSC population form more and larger neurospheres under hypoxia compared to normoxia. Therefore, hypoxia confers self-renewal capacities to non-CSCs as they are able to form spheres, and maintains the self-renewal of newly formed CSCs, which leads to larger sphere formation. In addition, the expression of pluripotency factors and HIF-2 α is increased in non-CSCs cultured in hypoxic conditions ([Heddleston et al., 2009](#)).

The hypoxia-induced dedifferentiation was also reported in pancreatic cancer ([Mu et al., 2021](#); [Zhu et al., 2013](#)). The intermittent hypoxic environment exposure of non-CSCs was able to induce

self-renewal capacities and high levels of HIF-1 α , light chain 3-II (LC3-II), and Beclin RNAs and proteins, which are involved in autophagy. The knock-down of HIF-1 α abolished the hypoxia-induced expression of LC3-II and Beclin, meaning that HIF-1 α is required for hypoxia-induced autophagy in CSCs. And, the inhibition of both HIF-1 α and autophagy reduced the hypoxia-induced CSC percentage, indicating that the HIF-1 α -mediated autophagy triggered by intermittent hypoxia contributes to the non-CSC dedifferentiation into CSC in pancreatic cancer (Zhu *et al.*, 2013).

The dedifferentiation of non-CSC populations from glioblastoma, hepatocellular carcinoma, and lung cancer cell lines under hypoxia were studied by Wang *et al.* (2017d). The non-CSC populations cultured in hypoxic conditions showed significantly higher proportions of CSCs and formed more spheres than those cultured in normoxic conditions, for the three cell lines. Hence, they demonstrated a hypoxia-induced formation of new CSCs from non-CSCs occurring in glioblastoma, hepatocellular carcinoma, and lung cancer. Moreover, immunofluorescence analysis of hypoxia-derived glioblastoma spheres revealed a high expression of the stem cell markers SOX2, OCT4, KLF4, NANOG, LIN28, CD133, CD15, and NESTIN. The culture of these spheres in a stem cell medium confirmed their self-renewal ability and extensive proliferation, while cultured in a differentiation medium, they showed adherent growth and morphology changes indicating their capability to perform asymmetric division. Furthermore, Wang *et al.* (2017c) found a high expression of HIF-1 α in the hypoxia-treated neurospheres as well as in tumor samples from mice raised in hypoxia. The silencing of HIF-1 α reduced the neurosphere formation rate and the CSC proportion in the hypoxia-treated non-CSC population, and impaired the tumorigenic capacities of cells as tumors in mice grew smaller compared to the control group (Wang *et al.*, 2017c). Then, Wang *et al.* (2022) confirmed that the HIF-1 α /HIF-2 α -SOX2 network induced the dedifferentiation of glioblastoma non-CSCs into CSCs under hypoxic conditions.

2.3.2 EXTRACELLULAR VESICLES

Extracellular vesicles (EVs) mediate the intercellular communication by transmitting a molecular content composed of several proteins, lipids, and nucleic acids. Additional details about EVs are given in Appendix 2 “Extracellular vesicles and dedifferentiation into cancer stem cells” at page 319.

The radio-induced EVs and chemo-induced EVs promote stem cell characteristics, amplifying the therapeutic resistance (Chung *et al.*, 2021; Kyjacova *et al.*, 2015; Meldolesi, 2022; Ramakrishnan *et al.*, 2020; Shen *et al.*, 2019; Yang *et al.*, 2021). Following radiotherapy, secreted EVs participate in the radiation-induced bystander effect, meaning that even unirradiated cells can be affected by radiotherapy effects through the uptake of EVs derived from irradiated neighboring ones (Jabbari *et al.*, 2020; Jella *et al.*, 2014; Szatmári *et al.*, 2017; Xu *et al.*, 2016). Moreover, the radiation treatment induces changes in the secreted EVs quantity and content, which can confer radiation resistance properties to recipient cells (Al-Mayah *et al.*, 2012; Jelonek *et al.*, 2016; Mutschelknaus *et al.*, 2016). For instance, in hepatocellular carcinoma, exosomes released by CSCs induces the expression of

NANOG in non-CSCs and enhance their drug resistance (Huang *et al.*, 2021).

Additionally, EVs derived from stromal cells also participate in the promotion of stem cell features and in the radio- or chemo-resistance of cancer cells. CAFs-derived EVs contribute to the maintenance and expansion of colorectal cancer CSCs through activation of stemness related pathways, such as the Wnt/ β -catenin signaling or TGF- β signaling, mediated by the EV-transported miRNAs (e.g. miR-92a-3p) or lncRNAs (e.g. H19) (Conigliaro *et al.*, 2015; Hu *et al.*, 2019a; Hu *et al.*, 2015; Liu *et al.*, 2020b; Ren *et al.*, 2018). In breast cancer, EVs from stromal cells containing three miRNAs (miR-21, miR-378e, and miR-143) increase mammospheres formation and promote the EMT and stemness phenotypes (Donnarumma *et al.*, 2017), and they can activate the STAT1 and Notch signaling promoting stemness and resistance to therapy (Boelens *et al.*, 2014).

The interconversion between CSCs and non-CSCs is mediated by EVs (Figure 14). Indeed, in their cargo, stemness regulatory molecules can induce a phenotypic switch of recipient cells, to maintain the balance between CSC and non-CSC populations or expand the CSC pool in response to therapy for instance (Sun *et al.*, 2018; Xu *et al.*, 2018).

First, to study the effect of CSC-derived exosomes on non-CSCs, Sun *et al.* (2020) used three glioblastoma non-CSC cell lines treated with exosomes from a CSC-like cell line cultured in a stem cell medium and showing high expression of CSC markers. Although the CSC and non-CSC populations were not purified using CSC markers, they reported an increase in tumorigenicity in the treated cells and enrichment of CSC-derived exosomes in Notch1 proteins. As the inhibition of Notch1 decreased the stemness features induced by CSC-exosomes in non-CSC cell lines, it revealed that the Notch signaling contributes to the exosome-mediated enhanced stemness of glioblastoma cells.

In colorectal cancer, Hu *et al.* (2019b) demonstrated the induced dedifferentiation of purified non-CSCs by CAFs-derived EVs during chemotherapy treatment. The non-CSC population treated with fibroblast-derived conditioned media showed higher sphere formation and higher CSC percentage upon chemotherapy treatment. Furthermore, the fibroblast-derived exosomes coupled with chemotherapy-induced more sphere generation in the non-CSC population compared to control cells, while the exosome-depleted conditioned media treatment-induced less sphere generation, indicating that the dedifferentiation is mediated by exosomes. They also found the presence of several Wnt ligands in the fibroblast-secreted vesicles, activating the Wnt/ β -catenin pathway in recipient cells. Overexpression of WNT3A in CAFs revealed that CAFs-derived EVs increase the nuclear β -catenin levels in treated non-CSCs and enhance their tumorigenicity *in vivo*.

2.3.3 INFLAMMATION AND SECRETED MOLECULES

Cytokines and other molecules that are secreted by a tumor or stromal cells act as paracrine communication between cells, activating signaling pathways related to stemness to promote the acquisition of a stem phenotype in non-CSCs (Figure 14) (Das *et al.*, 2020; Wijaya, 2019).

Pro-inflammatory mediators such as tumor necrosis factor α (TNF α) and IL-6 have been reported to induce the dedifferentiation toward a CSC state. Indeed, the IL-6 is known to induce an enrichment in CSCs through the non-CSC-to-CSC conversion in several cancer models including lung cancer, breast cancer (Arnold *et al.*, 2020b; Korkaya *et al.*, 2012; Liu *et al.*, 2014; Rodrigues *et al.*, 2018; Saha *et al.*, 2016; Vaziri *et al.*, 2021).

On one hand, to initiate the reacquisition of stem cell features such as the reexpression of CSC markers and self-renewal capacities, the IL-6 inflammatory signal activates the NF- κ B signaling (Korkaya *et al.*, 2012; Schwitalla *et al.*, 2013; Wang *et al.*, 2014a). Iliopoulos *et al.* (2009) highlighted the IL-6/NF- κ B signaling regulates the LIN28 expression via the STAT3 transcription factor activity. Then, Iliopoulos *et al.* (2011) reported that the IL-6 secretion mediates the balance between the CSC and non-CSC phenotypes within the breast cancer cell population. These results were confirmed by Saha *et al.* (2016) as they demonstrated the chemo-induced activation of the IL-6/NF- κ B loop contributes to the non-CSC-to-CSC conversion in breast cancer.

On the other hand, the IL-6 signal has been found to trigger the activation of the JAK/STAT signaling, also mediating the acquisition of stem cell properties (Jin, 2020). In breast cancer, the secreted IL-6 activates the JAK1-STAT3 signal transduction pathway in non-CSCs, up-regulating the OCT4 expression and triggering the non-CSCs conversion into CSCs (Kim *et al.*, 2013b). The IL-6 also activates the JAK2-STAT3 signaling to induce the non-CSCs dedifferentiation (Arnold *et al.*, 2020a; Liu *et al.*, 2014).

The TNF α produced by T-cells have been reported to promote the dedifferentiation of melanoma cells (Landsberg *et al.*, 2012).

The TGF- β secreted in the tumor microenvironment by stromal or cancer cells is also involved in the dedifferentiation of osteosarcoma, lung cancer, breast cancer, and colorectal cancer cells (Andriani *et al.*, 2016; Chaffer *et al.*, 2013; Nakano *et al.*, 2018; Zhang *et al.*, 2013a). The hypoxia-induced secretion of TGF- β 1 participates in the acquisition of stemness characteristics in osteosarcoma non-CSCs (Zhang *et al.*, 2013a). The TGF- β stimulation induces the TWIST1 and ZEB1 expression leading to the conversion of non-CSC into CSCs, in colorectal and breast cancer respectively (Chaffer *et al.*, 2013; Nakano *et al.*, 2018). Lastly, the sorted non-CSC population of lung cancer cells showed an increase in CSCs after TGF- β 1 treatment (Andriani *et al.*, 2016).

Finally, secreted chemokines, or chemotactic cytokines (C-X-C motif chemokine ligand (CXCL) and C-C motif chemokine ligand (CCL)), can also activate these previously mentioned signaling pathways such as the JAK/STAT one, or others such as the phosphatidylinositol 3-kinase (PI3K)/Akt one, which are also implicated in the CSC state (Chang *et al.*, 2013; Hernández-Vargas *et al.*, 2011; Korkaya *et al.*, 2009). In particular, CXCL12, also known as stromal cell-derived factor-1 (SDF-1), and its C-X-C motif chemokine receptors (CXCRs) 4 and 7 (CXCR4 and CXCR7) have been associated with CSC enrichment (Dubrovskaya *et al.*, 2012; Kong *et al.*, 2016; Tang *et al.*, 2016; Zhang *et al.*,

2013b). Additionally, previous work in the laboratory revealed that the CXCL1 and CCL5 chemokines were highly secreted after irradiation of SUM159PT breast cancer cells, and their inhibition or the neutralization of their receptors combined with radiotherapy resulted in decreased tumorigenic capacities (Bailleul, 2018).

2.3.4 EPIGENETICS

The reversibility of epigenetic modifications enables the phenotypic plasticity of cancer cells. Indeed, interactions of the cell with environmental factors generate epigenetic modifications that result in changes in gene expression in response to the stimuli (Hass *et al.*, 2020). Hence, the re-expression of pluripotency factors and self-renewal related genes can be unlocked through epigenetic changes and contribute to the phenotypic switch toward the stem cell state (Figure 14) (French and Pauklin, 2021; Markopoulos *et al.*, 2019; Muñoz *et al.*, 2012; Poli *et al.*, 2018; Safa *et al.*, 2015; Wainwright and Scaffidi, 2017).

NON-CODING RNAs

As previously stated, numerous miRNAs are found to be involved in the acquisition of a stem cell phenotype, *via* their transportation in EVs or through the activation of signaling pathways regulating miRNAs expression, and lncRNAs also participate in the induction of the stem cell state.

The lncRNA H19 along with the let-7 miRNAs have a role in the stem cell phenotype induction as they are known to be pluripotency regulators (Albino *et al.*, 2016; Jiang *et al.*, 2016; Lecerf *et al.*, 2020; Ma *et al.*, 2021; Peter, 2009; Ren *et al.*, 2018). H19 is overexpressed in CSCs and regulates with the let-7 family of miRNAs the expression of the pluripotency factor LIN28. Indeed, the H19 negatively regulates the let-7 miRNAs, which are negative regulators of the LIN28 gene. This negative feedback loop leading to the expression of LIN28 is a marker of pluripotency involved in the CSC maintenance and expansion, while the let-7 miRNAs, negative regulators of this loop, are associated with a differentiated phenotype (Albino *et al.*, 2016; Cai *et al.*, 2013; Yang *et al.*, 2010; Yu *et al.*, 2007a). Guo *et al.* (2019) have reported that the let-7/LIN28 axis is involved in the generation of breast CSCs. Especially since the let-7/LIN28 axis involves the IL6 and the NF- κ B signaling, known regulators of the dedifferentiation, both H19 and let-7 miRNAs may participate in the non-CSC-to-CSC conversion (Iliopoulos *et al.*, 2009; Iliopoulos *et al.*, 2011).

Many other miRNAs are involved in the acquisition of stemness, including the miR-200 family. The miR-200 family of miRNAs is widely known to regulate the EMT *via* the ZEB family transcription factor negative feedback loop (Andriani *et al.*, 2016; Cano and Nieto, 2008; Korpál *et al.*, 2008). Yet, they have also been described as mediating the stem cell phenotype (Peter, 2009; Wellner *et al.*, 2009). For instance, the miR-200a negatively regulates the CSC state in pancreatic and liver cancer (Lu *et al.*, 2014b; Wang *et al.*, 2015b), and the miR-200c in breast cancer (Feng *et al.*, 2015b; Liu *et al.*, 2018a; Shimono *et al.*, 2009). The miR-200b downregulates the generation and mainte-

nance of CSC *via* the inhibition of Suz12 (a component of the PRC2 complex), inducing the loss of the H3K27me3-mediated repression of the E-cadherin gene (CDH1) in breast cancer (Iliopoulos *et al.*, 2010). Additionally, Lim *et al.* (2013) found that miR-200s are involved in the plasticity between non-stem and stem-like phenotypes in immortalized mammary cells and that their loci have different DNA methylation and histones modification profiles between the two phenotypes. Indeed, the repressive methylated state of the miR-200 loci is maintained by the miR-22-mediated inhibition of TET demethylase in breast CSCs (Song *et al.*, 2013b). Finally, Chaffer *et al.* (2013) demonstrated that the miR-200 regulation of ZEB1 plays a key role in the non-CSC-to-CSC conversion in breast cancer, as the miR-200 inhibition severely impairs their dedifferentiation rate.

Lastly, lincRNA-ROR is a lincRNA known to be involved in the reprogramming into iPS cells, and in the CSC phenotype in glioblastoma and pancreatic cancer (Feng *et al.*, 2015a; Fu *et al.*, 2017; Loewer *et al.*, 2010; Zhan *et al.*, 2016). And, the miR-21 is also a well-known regulator of both EMT and CSC phenotype and could participate in the dedifferentiation (Han *et al.*, 2012a; Han *et al.*, 2012b; Ni *et al.*, 2018; Sekar *et al.*, 2016).

DNA METHYLATION

The induced reprogramming of somatic cells into iPS cells is accompanied by a reset of epigenetic marks, especially by global DNA demethylation (Bhutani *et al.*, 2009; Costa *et al.*, 2013; Gao *et al.*, 2013; He *et al.*, 2017; Park *et al.*, 2020). The presence of DNA methylation at gene promoter is generally associated with its transcriptional repression. Details about DNA methylation and epigenetic reprogramming are given in chapter 4 “DNA methylation: principles and methodology” at page 105. The involvement of ten-eleven translocation (TET) 1 in the global DNA demethylation during the establishment of a stem-like pluripotent state indicates that TET enzymes may be also implicated in the generation of CSCs, through the epigenetic reactivation of previously silenced genes (Costa *et al.*, 2013; French and Pauklin, 2021; Gao *et al.*, 2013; Olariu *et al.*, 2016; Stricker and Pollard, 2014). Globally, the CSCs have different DNA methylation signatures compared to non-CSC, demonstrating the importance of DNA methylation changes to enable the phenotypic plasticity between the non-CSC and CSC states (Helou *et al.*, 2014; Lee *et al.*, 2015). The DNA methylation-associated enzymes are therefore required to apply the DNA methylation changes occurring through the phenotypic switch, but their role remains mostly unclear. In general, DNMT1 has been found to promote the formation of CSCs (Liu *et al.*, 2014; Pathania *et al.*, 2015; Zagorac *et al.*, 2016), and DNMT3B has been reported as suppressed by the miR-221, associated with stemness in breast cancer (Roscinno *et al.*, 2015).

The DNA methylation or demethylation of specific loci contributes to phenotypic changes. For instance, the promoter of NANOG is found hypomethylated in hepatocellular carcinoma and liver cancer CSCs (Liu *et al.*, 2020c; Wang *et al.*, 2013c). In liver cancer, Liu *et al.* (2020c) reported differences in DNA methylation at the NANOG gene promoter and DNMT1 expression levels, between

non-CSCs and CSCs. They found that the miR-135a downregulates the DNMT1 expression, reducing the DNA methylation of the NANOG promoter and resulting in the SMYD4-mediated activation of its transcription, to enable the acquisition of stem cell properties and dedifferentiation into CSCs. The promoter of the CD133 gene, a CSC marker commonly used in different cancer models, is found differentially methylated in CSCs in several cancers, such as breast cancer, glioblastoma, hepatocellular carcinoma, and ovarian cancer (Baba *et al.*, 2009; Kagara *et al.*, 2012; Yi *et al.*, 2008; You *et al.*, 2010). Several genes are found regulated by DNA methylation in CSCs, including the GATA6 gene, a transcription factor associated to differentiation and found hypermethylated in glioblastoma CSCs, and SOX9 gene promoting self-renewal via the asymmetrical-to-symmetrical cell division switch and found hypomethylated in pancreatic ductal adenocarcinoma CSC (Lee *et al.*, 2015; Liu *et al.*, 2016; Sun *et al.*, 2013; Sun and Yan, 2020). These data suggest that the methylation level of specific localization might be more important in the determination of the cell phenotype than the global DNA pattern (French and Pauklin, 2021).

HISTONE MODIFICATIONS

The histone post-translational modifications (PTMs) such as methylation or acetylation of lysine residues at their tails, contribute to the establishment of the chromatin state and subsequent gene transcription or gene silencing. Several enzymes, including histone methylase (HMT) and histone deacetylase (HDAC) along with key co-factors, can modify the histones marks, resulting in chromatin condensation or decondensation and thus changes in gene expression. In general, the acetylation of histones is correlated with transcriptional activity, as well as for the H3K4me3 mark, while the H3K27me3 is associated to gene silencing.

The presence of two different histone marks with opposing effects at the same localization is called bivalency or poised chromatin. This bivalency is found generally at promoter regions of genes important for the development. Indeed, the presence of both activating or silencing marks, such as H3K4me3-H3K27me3 marks, keep the gene silenced but allow its rapid transcriptional activation upon cell fate commitment (Bapat, 2013; Bernstein *et al.*, 2006; Easwaran *et al.*, 2014). As the interconversion between CSC and non-CSC states depends on the capacity to activate or repress transcriptional programs, the poised chromatin may have a central role in the CSC plasticity (Chaffer *et al.*, 2013; Suvà *et al.*, 2013; Wainwright and Scaffidi, 2017).

In acute myeloid leukemia (AML) and glioblastoma, the CSCs and non-CSCs have different histone modifications profiles (Rheinbay *et al.*, 2013; Yamazaki *et al.*, 2012). The bivalency of promoters in AML CSCs is lost when they differentiate into progenitors, meaning that the reverse process could contribute to the dedifferentiation of CSCs (Yamazaki *et al.*, 2012). In particular, Chaffer *et al.* (2013) found that the promoter of the ZEB1 gene, which mediates the dedifferentiation of breast cancer cells, exhibits bivalent chromatin patterns (H3K4me3 and H3K27me3 marks) in non-CSCs but only activation marks in CSCs (H3K4me3 and H3K79me2 marks). Hence, it indicates that the loss of

repressive marks at bivalent promoters can enable the reexpression of stemness-related genes for the cell to transit to a CSC phenotype.

To modify the histones marks through the phenotypic switches, the activity of histones modifying enzymes is required. The involvement of histone acetylation in the non-CSC-to-CSC conversion was demonstrated by [Debeb *et al.* \(2012\)](#) as HDAC inhibitors treatment induced the dedifferentiation of breast CSCs through the up-regulation of the Wnt/ β -catenin pathway, as explained in section 2.2.2 “HDAC inhibitors induced dedifferentiation” at page 75. In lung cancer, [Saijo *et al.* \(2016\)](#) showed that oxidative stress induces the repression of the HDAC8. Thus, the presence of histone acetylation modifications at the HOX5A and SOX2 promoters enables the SOX2 expression *via* the HOX5A transcription activity and leads to the acquisition of a stem cell phenotype in lung cancer cells. Hence, it also explains the generation of CSCs after HDAC inhibitors exposure as it triggers the expression of HOX5A and SOX2.

The polycomb repressive complexes (PRC) 1 and 2 (responsible for H2AK119Ub1 and H3K27me3 marks) are strongly linked to the maintenance of pluripotency in several cancers ([Abdouh *et al.*, 2009](#); [Gorodetska *et al.*, 2019](#); [Orzan *et al.*, 2011](#); [Proctor *et al.*, 2013](#); [Safa *et al.*, 2015](#); [Suvà *et al.*, 2009](#); [Vlerken *et al.*, 2013](#); [Wen *et al.*, 2015](#)). EZH2, the catalytic component of PRC2 mediating the H3K27me3, participates in the CSC plasticity in glioblastoma through the regulation of NANOG, WNT1, and BMP5 genes and is also known to regulate the expression of the GATA6 gene, coding for transcription factor associated to cell lineage decision and differentiation, in several cancer types ([Natsume *et al.*, 2013](#); [Patil *et al.*, 2020](#); [Tan *et al.*, 2019](#); [Zang *et al.*, 2020](#)). Altogether, the activity of these histone modifiers could play a key role in the dedifferentiation of non-CSCs into CSCs.

KEY POINTS

- ➔ The non-CSC-to-CSC conversion is regulated at several levels by different machineries and mediators, from the microenvironment composition and signals to the modulation of gene transcription by epigenetic mechanisms.
- ➔ An hypoxic environment, caused by chemotherapy or radiotherapy for example, leads to the emergence of CSCs in the non-CSC population through increased levels of hypoxia-inducible factors (HIFs), in several cancer models, especially in glioblastoma.
- ➔ The intercellular communication *via* extracellular vesicles (EVs) also participates in the acquisition of a CSC phenotype. In the context of a radiotherapy treatment, EVs contribute to the radiation-induced bystander effect, affecting the phenotype of unirradiated recipient cells. Further information linking EVs and dedifferentiation is available in Appendix 2 “Extracellular vesicles and dedifferentiation into cancer stem cells” at page 319.
- ➔ Secreted pro-inflammatory factors (e.g. $\text{TNF}\alpha$, interleukins, cytokines) can activate signaling pathways (e.g. $\text{NF-}\kappa\text{B}$, JAK/STAT), inducing the dedifferentiation of non-CSCs into CSCs.
- ➔ Epigenetic mechanisms, non-coding RNAs, DNA methylation, and histones post-translational modifications, contributes as well to the phenotypic conversion. For instance, TET enzymes and the global DNA demethylation may be involved in the epigenetic reprogramming enabling the acquisition of stem cell features. Additionally, the bivalent (or poised) chromatin, characterized by the presence of opposing epigenetic marks at the same loci, can have a central role in the interconversion between the CSC and non-CSC states, as the lost of the repressive marks would rapidly enable the reexpression of key genes, allowing the reacquisition of stem cell properties.

[^ Back to Table of Contents](#)

3

BREAST CANCER AND BREAST CANCER STEM CELLS

☰ CHAPTER CONTENTS

3.1	THE MAMMARY GLAND	87
3.1.1	DESCRIPTION	87
3.1.2	MAMMARY STEM CELLS	89
3.2	BREAST CANCER	90
3.2.1	INDICATORS AND STATISTICS	90
3.2.2	TUMORIGENESIS	92
3.2.3	TUMORAL PROGRESSION	94
3.2.4	HETEROGENEITY AND CLASSIFICATIONS	94
3.2.5	THERAPIES	99
3.3	BREAST CANCER STEM CELLS	101
3.3.1	MARKERS AND ISOLATION	101
3.3.2	FUNCTIONAL TESTS	103
	📌 KEY POINTS	104

[^ Back to Table of Contents](#)

3.1 THE MAMMARY GLAND

3.1.1 DESCRIPTION

The breast is a mammalian organ that has the function of maternal milk secretion after child-birth to feed the newborn. It is an exocrine mammary tubuloalveolar gland surrounded by fatty and connective tissue (Figure 15). Its structure is supported by the pectoral muscles, and its integrity is ensured by a connective tissue composed of the fibroglandular ridges of Duret and Cooper's ligaments.

From 15 to 25 irregular lobes compose the mammary gland, connected to a milk duct for milk transportation to the nipple. The milk duct gradually branches out into interlobular ducts, each leading to a lobule composed of numerous acini. An acinus is composed of a lumen surrounded by lactocytic epithelial cells that secrete milk, overlaid by myoepithelial cells. The acini are connected to the intralobular duct, which joins the interlobular duct (Figure 15). The myoepithelial cell contraction

during lactation allows the ejection of milk towards ducts (Hassiotou and Geddes, 2013; Pinamonti *et al.*, 2017).

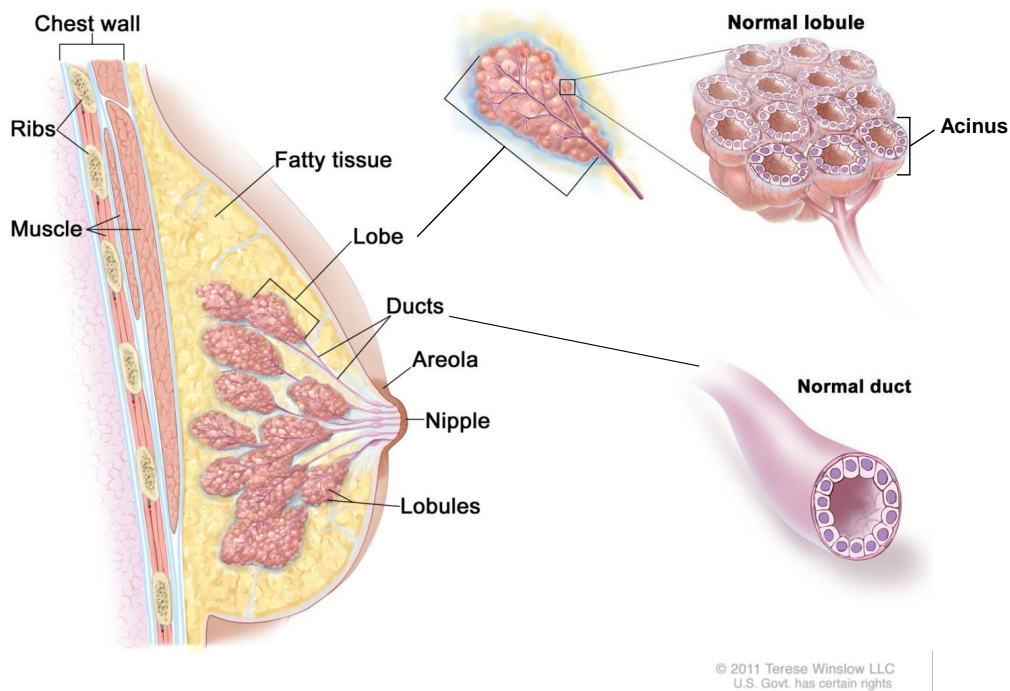


FIGURE 15 ANATOMY OF THE FEMALE MAMMARY GLAND. Representation of a sagittal section of the mammary gland in women. The mammary gland is composed of glandular and adipose, or fatty, tissues, supported by ligaments. A ductal system drains the secretory tissue to store and transport milk to the nipple during lactation. The glandular tissue is composed of 15-25 lobes that comprise 20-40 lobules containing 10–100 acini in which the milk is produced. Figure adapted from the [Medical And Scientific Illustration, Breast, Terese Winslow LLC website](#), © Terese Winslow LLC for the National Cancer Institute.

The mammary epithelium develops at puberty in women under hormonal stimulation. During gestation, high concentrations of progesterone, estrogen, and placental lactogenic hormone activate cellular proliferation, through the growth factors production, leading to extensive branching of the ductal tree, and acini differentiation in preparation for milk production. After childbirth, modification of the hormone balance between progesterone and prolactin induces milk secretion. Finally, at the cessation of lactation, the mammary gland progressively involutes to return to its pre-pregnant state *via* epithelial and myoepithelial cells apoptosis and proteinase-driven tissue remodeling (Fu *et al.*, 2014).

The morphological and functional unit of the breast called the **terminal duct lobular unit (TDLU)** is the structure formed by an extralobular terminal duct and a lobule (acini connected to the intralobular duct). Histologically, the TDLU is composed of two main cell types: an inner luminal layer of cylindrical or cubic glandular epithelial cells also called ductal or alveolar luminal cells, and an outer basal layer composed of myoepithelial cells (Figure 16) (Fu *et al.*, 2020; Pinamonti *et al.*, 2017; Visvader, 2009).

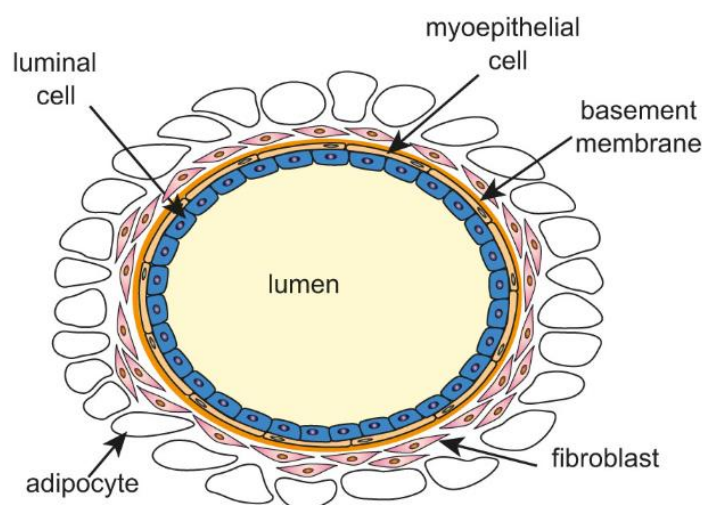


FIGURE 16 STRUCTURE AND CELL TYPE COMPOSITION OF THE MAMMARY GLAND. Representation of the mammary gland ductal structure and cell type composition. Two types of luminal cells are found depending on their localization, ductal ones in ducts and alveolar ones in alveoli (acini). Figure from [Fu et al. \(2020\)](#).

3.1.2 MAMMARY STEM CELLS

The breast development during puberty and regeneration through successive cycles of pregnancy suggested repopulation and renewal capacities provided by **tissue-resident adult stem cells**. The presence of stem cells within the mammary gland has been first implied by [DeOme et al. \(1959\)](#) as the transplantation of breast tissue fragments are able to reform ductal outgrowths resembling the normal mammary epithelial tree, and was supported later by the work of [Daniel et al. \(1968\)](#) demonstrating their self-renewal capacity through serial transplantation. However, **mammary stem cells (MaSCs)** were identified and isolated for the first time by [Shackleton et al. \(2006\)](#). They demonstrated that a single MaSC can regenerate an entire structure of the mammary gland upon transplantation. This repopulating activity into different mammary cell subtypes demonstrates their multidifferentiative and self-renewal capabilities and thus, they are characterized as stem cells.

Several markers are used for their isolation, in mice the $\text{Lin}^- \text{CD24}^+ \text{CD29}^{\text{high}}$ (Lin = lineage, CD29 = integrin $\beta 1$) cells are generally referred to murine MaSCs ([Fu et al., 2020](#); [Shackleton et al., 2006](#)). The isolation of human MaSCs in several studies relies on several combinations of markers to define the population, such as $\text{Lin}^- \text{CD49f}^+ \text{EpCAM}^{-/\text{low}}$ (CD49f = integrin $\alpha 6$, EpCAM = epithelial cell adhesion molecule), $\text{CD24}^{\text{high}} \text{CD49f}^{\text{high}} \text{DNER}^{\text{high}}$ (DNER = delta/notch like EGF repeat containing) and $\text{CD10}^+ \text{EpCAM}^-$ (CD10 = membrane metalloendopeptidase) ([Bachelard-Cascales et al., 2010](#); [Fu et al., 2020](#); [Pece et al., 2010](#); [Shehata et al., 2012](#)).

Hence, **the mammary epithelium is organized hierarchically** with stem cells and progenitors giving rise to luminal and myoepithelial lineages, which is necessary for the mammary gland development during puberty and gestation, and for the tissue homeostasis ([Figure 17](#)) ([Fu et al.,](#)

2020; Visvader, 2009). The MaSCs are qualified as multipotent or bipotent stem cells, as they differentiate into the two primary epithelial cell lineages, luminal and myoepithelial, but the luminal lineage can be further subdivided into the ductal and alveolar sublineages. Indeed, a subset of cycling ER+ progenitors produces luminal ER+ cells to line the ducts. Yet, through transplantation and lineage tracing studies, controversy remains about whether or not there is an intermediate bipotent progenitor state between the MaSCs and luminal/basal unipotent progenitors in the adult (Fu *et al.*, 2020; Keymeulen *et al.*, 2011; Rios *et al.*, 2014; Visvader, 2009).

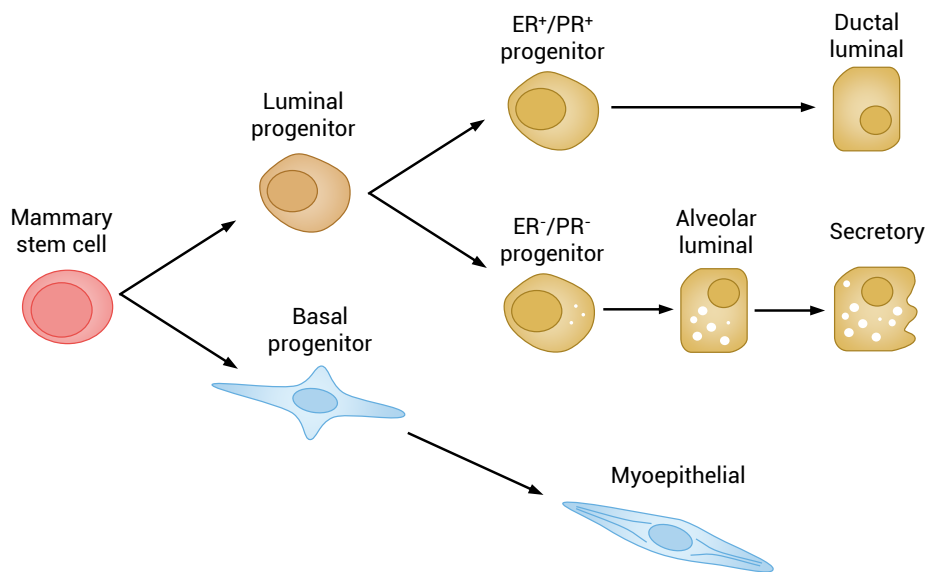


FIGURE 17 MAMMARY EPITHELIAL DIFFERENTIATION HIERARCHY. The mammary epithelium is organized in a hierarchy, the multipotent/bipotent MaSCs give rise to unipotent luminal and basal progenitors. The luminal lineage is subdivided into two sublineages, the ductal luminal one and the alveolar luminal one. Under hormonal stimuli at the end of pregnancy, the alveolar luminal cells can further differentiate to secrete milk. The basal progenitors give rise to myoepithelial cells forming the contractile tissue necessary for milk expulsion. Figure adapted from Fu *et al.* (2020).

3.2 BREAST CANCER

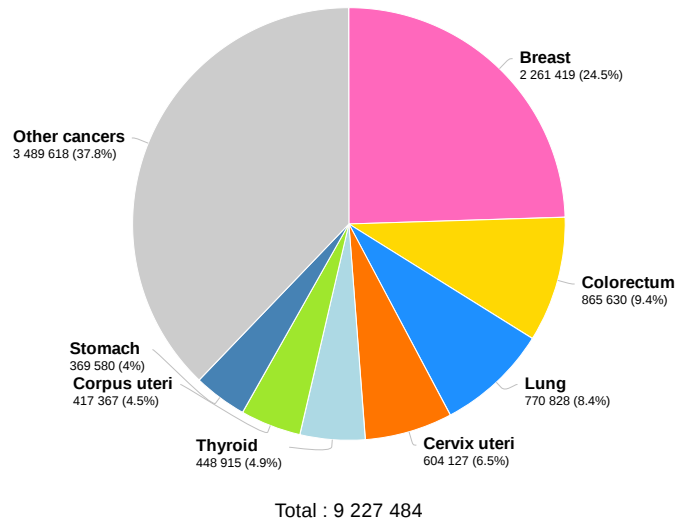
3.2.1 INDICATORS AND STATISTICS

WORLDWIDE

In 2020, 2.26 million new cases of breast cancer and 685,000 deaths due to breast cancer have been reported worldwide by the GLOBOSCAN project (Sung *et al.*, 2021). Breast cancer is the most diagnosed cancer as it represents 11.7% of all diagnosed cancers, for both sexes (Figure 2A). It is the fourth cause of cancer mortality as it causes 6.9% of all cancer deaths, for both sexes (Figure 2B). The 5-year prevalence, meaning the number of people alive within 5 years after a breast cancer diagnosis, is estimated at 7.79 million (17.7% of all cancers, for both sexes). It is estimated that nearly 30% of patients with breast cancer, who became disease-free after being treated,

A

Estimated number of new cases in 2020, worldwide, females, all ages

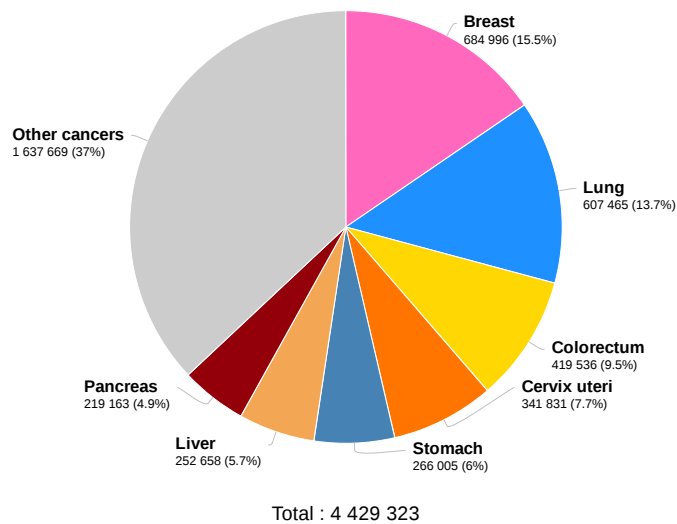


Data source: Globocan 2020
Graph production: Global Cancer Observatory (<http://gco.iarc.fr>)

International Agency for Research on Cancer
World Health Organization

B

Estimated number of deaths in 2020, worldwide, females, all ages



Data source: Globocan 2020
Graph production: Global Cancer Observatory (<http://gco.iarc.fr>)

International Agency for Research on Cancer
World Health Organization

FIGURE 18 ESTIMATED NEW CASES AND DEATHS BY CANCER SITES FOR WOMEN IN 2020. **A.** Estimated number of new cases (incidence) in 2020, worldwide, for women and all ages. **B.** Estimated number of deaths (mortality) in 2020, worldwide, for women and all ages. Graph and data from GLOBOSCAN 2020, IARC, WHO, available at the [Cancer Today - Global Cancer Observatory website](http://cancer.gov).

experience disease recurrence ([Colleoni et al., 2016](#)).

For women, breast cancer is the most diagnosed cancer, representing 24.5% of all new cancer cases ([Figure 18A](#)), and the first cause of cancer mortality, causing 15.5% of all cancer deaths ([Figure 18B](#)). As the estimated 5-year prevalence of breast cancer among women represent 33.7% of all cancers, it means that one-third of the women alive within 5 years after the cancer diagnosis have been diagnosed with breast cancer.

IN FRANCE

In France, 58,500 new cases of breast cancer and 12,100 deaths due to breast cancer have been reported in 2018. The prevalence for 2017 is estimated at 913,089 people alive who have been diagnosed with cancer in France.

For women, breast cancer is also the most diagnosed cancer in France, representing 33% of all new cancer cases, and the first cause of cancer mortality, causing 17.8% of all cancers death, in 2018 ([The French National Cancer Institute \(INCa\) website](#); [Panorama des cancers en France, 2021](#)).

3.2.2 TUMORIGENESIS

Cancer is a disease characterized by the presence of abnormal cells proliferating excessively and anarchically, uncoordinated with other tissues, and unresponsive to physiological regulatory mechanisms. The transition of a normal cell to a neoplastic cell able to form malignant tumors requires the acquisition of functional capabilities, defined and named as the **“Hallmarks of Cancer”** by Hanahan and Weinberg in [2000](#), [2011](#), and [2022](#) ([Figure 12A](#)). Indeed, the initial six hallmarks were complemented twice with emerging hallmarks and enabling characteristics that facilitate the acquisition of hallmark capacities.

Hallmarks of cancer:

- Sustaining proliferative signaling (2000)
- Evading growth suppressors (2000)
- Resisting cell death (2000)
- Enabling replicative immortality (2000)
- Inducing or accessing vasculature (2000)
- Activating invasion & metastasis (2000)
- Deregulating cellular metabolism (2011)
- Avoiding immune destruction (2011)
- Unlocking phenotypic plasticity (2022)

Enabling characteristics:

- Genome instability & mutation (2011)
- Tumor-promoting inflammation (2011)
- Nonmutational epigenetic reprogramming (2022)
- Polymorphic microbiomes (2022)
- Senescent cells (functionally important cell type, 2022)

Thus, the **emergence of breast cancer** is due to the acquisition of pro-tumor characteristics, or hallmarks, in luminal or myoepithelial cells, caused by genomic instability, epigenetic changes, and tumor microenvironment (Hanahan and Weinberg, 2011; Karsli-Ceppioglu *et al.*, 2014; Lee and Muller, 2010). Genetic alterations are DNA sequence modifications such as point mutations, amplification of the gene copy number, deletions, or chromosomal rearrangements, while epigenetic alterations modify the gene expression without DNA sequence changes. Additionally, the tumor microenvironment influence both genetic and epigenetic instability but also promotes tumorigenic signaling pathways enabling the tumor progression (Hanahan and Coussens, 2012; Hanahan and Weinberg, 2011).

Genes undergoing expression changes that cause the acquisition of hallmarks capacities are named **oncogenes**, promoting tumor growth, and **tumor suppressor genes**, inactivated as they repress tumor growth.

On one hand, the most overexpressed oncogenes in breast cancer are ERBB2 — commonly referred to as the human epidermal growth factor receptor 2 (HER2) —, MYC, and cyclin D1 (CCND1) (Lee and Muller, 2010; Osborne *et al.*, 2004). The HER2 gene is amplified and overexpressed in 20%-30% of invasive breast cancer, it encodes a transmembrane tyrosine kinase growth factor receptor that promotes proliferation, angiogenesis, cell motility, metastases, and resistance to apoptosis, *via* the activation of several pathways (mitogen-activated protein kinase (MAPK) signaling and PI3K/Akt signaling) (Slamon *et al.*, 1987; Vijver *et al.*, 1988; Yarden and Sliwkowski, 2001). The MYC gene encodes a nuclear phosphoprotein that participates in gene transcription regulation and is amplified and overexpressed in 15%-25% of breast cancers. MYC is associated with poor prognosis, aggressive clinical features, and resistance to chemotherapy, as it controls proliferation, differentiation, metabolism, apoptosis, stress pathways, and mechanisms of drug resistance (Fallah *et al.*, 2017; Varley *et al.*, 1987; Xu *et al.*, 2010). The CCND1 gene encodes the cyclin D1 protein which regulates the cell cycle through temporal coordination of cell division events, and it is overexpressed in 40%-50% of breast tumors and amplified in 10%-20% of cases (Roy and Thompson, 2006; Steeg and Zhou, 1998).

On the other hand, the loss of function of tumor suppressor genes is effective when both alleles are inactivated, thus it requires two genetic hits (Knudson, 2001). Hence, an hereditary transmitted mutation of one allele confers a genetic predisposition for cancer. The most known breast cancer

predisposition genes are breast cancer genes (BRCA1 and BRCA2), involved in the DNA damage response and DNA repair (Roy *et al.*, 2012). Mutations in either BRCA1 or BRCA2 are responsible for 80% of hereditary breast cancer cases (5-10% of breast cancer cases are inherited) (Godet and Gilkes, 2017; Rosen *et al.*, 2003). The silencing of BRCA1 can also be caused by epigenetic modifications such as its promoter hypermethylation (Pang *et al.*, 2012; Yamashita *et al.*, 2015). Additionally, the most-studied tumor suppressor gene P53 (or TP53) is found to be mutated in 20%-30% of breast carcinomas (Bertheau *et al.*, 2013; Hollstein *et al.*, 1991).

3.2.3 TUMORAL PROGRESSION

Within the mammary terminal duct lobular unit, cancer cells can arise from either the lobular (acini) or the ductal epithelium. Lobular carcinomas are less prevalent than ductal ones as they account for only 4%-10% of breast lesions biopsies (Donaldson *et al.*, 2018). The Figure 19 represents the stages of breast carcinoma progression from the ductal epithelium. The different stages of breast lesions are classified depending on their morphology and histology, based on biopsies, as followed:

- **Hyperplasia (H)** is characterized as an overgrowth of cells in the duct (ductal hyperplasia) or lobule (lobular hyperplasia) and remains benign.
- **Atypical hyperplasia (AH)** is characterized as an accumulation of monomorphic cells forming unusual architecturally complex patterns and is a high-risk benign lesion that can evolve in cancer (Hartmann *et al.*, 2015). Two types are distinguished, atypical ductal hyperplasia (ADH) and atypical lobular hyperplasia (ALH). The differentiation of atypical hyperplasia from low-grade carcinoma *in situ* rests on the absence of all the features of carcinoma *in situ* (Page and Rogers, 1992; Pinder and Ellis, 2003).
- **Carcinoma *in situ* (CIS)** is characterized as a more extensive accumulation of abnormal cells filling the entire lumen with no evidence of invasion across the basement membrane and is stage 0 of breast cancer (Gorringe and Fox, 2017; Pinder *et al.*, 2010; Tower *et al.*, 2019). Two types are distinguished, ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS).
- **Invasive carcinoma (IC)** is characterized by cancer cells breaking through the basement membrane and invading the surrounding stroma (Gannon *et al.*, 2014). Two types are distinguished, invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC).
- **Metastatic breast cancer** is characterized by the formation of distant tumors, named metastases. To form a metastasis, the cancer cells must enter the vasculature by intravasation, survive in the absence of adhesion, exit the blood system by extravasation, and regrow a new tumor in a different microenvironment (Vargo-Gogola and Rosen, 2007; Veer and Weigelt, 2003; Weigelt *et al.*, 2005). The preferential metastatic sites for breast cancer are the bone (65.1%), lung (31.4%), liver (26%), and brain (8.8%) (percentages from Chen *et al.*, 2017b).

3.2.4 HETEROGENEITY AND CLASSIFICATIONS

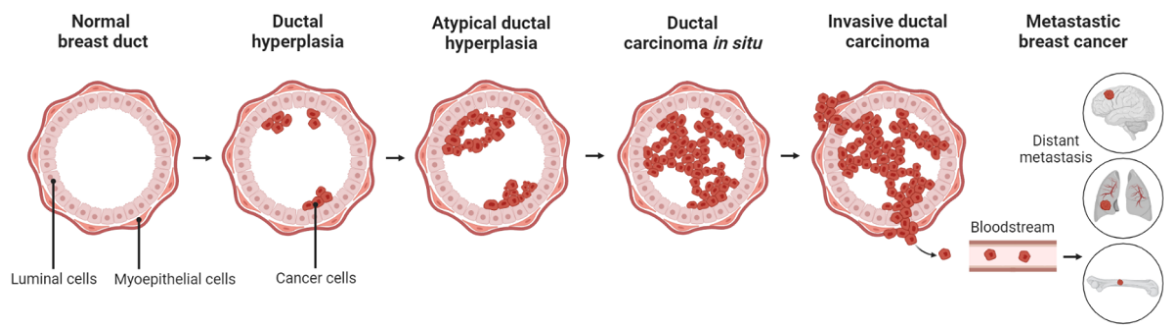


FIGURE 19 BREAST CANCER PROGRESSION. Representation of breast cancer progression stages. Atypical ductal hyperplasia (ADH) (or atypical lobular hyperplasia, ALH) and ductal carcinoma *in situ* (DCIS) (or lobular carcinoma *in situ*, LCIS) are pre-invasive lesions where cancer cells are confined in the duct. ADH progresses to DCIS at the complete filling of the mammary duct with cancer cells. The carcinoma becomes invasive (invasive ductal carcinoma, IDC, or invasive lobular carcinoma, ILC) when the myoepithelium and basement membrane are breached, allowing cancer cells to escape beyond the mammary duct confinement. The intravasation of cancer cells in the bloodstream followed by extravasation in a distant site lead to the formation of a distant metastasis. Created with BioRender.com

HISTOLOGICAL CLASSIFICATION

As previously detailed, breast cancers are classified based on histological and morphological criteria from biopsies, that distinguish breast cancer based on morphological and architectural features that define invasiveness and aggressiveness of the lesion. Two types of breast cancers are distinguished depending if the cancer cells have broken through the basement membrane or not: **carcinoma *in situ* (CIS)** and **invasive carcinoma (IC)**.

The CIS is subdivided depending on the epithelium of origin, giving **ductal carcinoma *in situ* (DCIS)** and **lobular carcinoma *in situ* (LCIS)**. The DCIS is further subdivided into **5 subgroups** based on their different architectures: the comedo, cribriform, micropapillary, papillary, and solid, and can be sub-classified into **3 grades**, low, intermediate, and high, depending on the lesion size and the presence of specific morphological features (Makki, 2015; Malhotra *et al.*, 2010).

Invasive carcinomas are also subdivided depending on the epithelium of origin, into **invasive ductal carcinoma (IDC)** (85%–95%) or **invasive lobular carcinoma (ILC)** (5%–15%). Several types of IDC and ILC have been defined depending on their morphology and architecture. **IDC subtypes:** no specific type (40%-75% of all invasive carcinomas), tubular carcinoma, invasive cribriform carcinoma, mucinous carcinoma, medullary carcinoma, invasive papillary carcinoma, invasive micropapillary carcinoma, apocrine carcinoma, neuroendocrine tumor, metaplastic carcinoma, lipid-rich carcinoma, secretory carcinoma, oncocytic carcinoma, adenoid cystic carcinoma, and acinic cell carcinoma. **ILC subtypes:** classic type, pleomorphic lobular carcinoma, histiocytoid carcinoma, signet ring carcinoma, and tubulolobular carcinoma (Makki, 2015).

HISTOLOGICAL GRADES

The **Nottingham grade (or score) system** is a histological grading of breast tumors based on their differentiation degree. It was originally developed by [Bloom and Richardson \(1957\)](#), and modified by [Elston and Ellis \(1991\)](#) from the hospital of Nottingham. Scores from 1 to 3 are determined based on tubules (group of cells forming a ring-like structure) proportions, nuclear pleomorphism (non-uniformity of nucleus shapes and sizes), and mitotic rate (number of mitotic figures per area) ([Table 2](#)) ([Elston and Ellis, 1991](#)). The addition of these scores gives the histological grade of the tumor from I to III, which serves for prognosis and clinical outcome prediction. A total score between 3 and 5 gives a **low grade** of I, the tumor is well differentiated with low proliferation. A total score of 6 or 7 gives an **intermediate grade** of II, the tumor is moderately differentiated. And, a total score of 8 or 9 gives a **high grade** of III, the tumor is poorly differentiated and tumor cells are highly proliferative.

HISTOLOGICAL FEATURES	SCORE = 1	SCORE = 2	SCORE = 3
Tubules	Majority (> 75%)	Moderate (10-75%)	Low (< 10%)
Nuclear pleomorphism	Small regular and uniform	Moderate increase in size and variability	Strong variability
Mitoses number per area*	0-9	10-19	>20

TABLE 2 HISTOLOGICAL GRADING SYSTEM OF BREAST CANCER. * 10 fields at the tumor periphery, number thresholds depends on the microscope/objective used ([Elston and Ellis, 1991](#)).

CLINICAL CLASSIFICATION

The **Tumor-Node-Metastasis (TNM) classification** is an anatomic classification of mammary tumors based on macroscopic features and tumor invasive capacity. The combination of 3 parameters: \textcircled{T} the primitive tumor size, \textcircled{N} the localization and invasion degree of lymph nodes, and \textcircled{M} the presence of metastasis, determines the clinical stages from 0 to IV, which help clinicians to predict the cancer prognosis and evolution to adapt therapies ([Table 3](#)) ([Brierley et al., 2016](#)).

BIOMARKERS

Three main biomarkers are commonly used to characterized mammary tumors: the **estrogen receptor (ER)**, the **progesterone receptor (PR)** and the **human epidermal growth factor receptor 2 (HER2)** (also known as ERBB2). The **ER- and PR-receptor statuses** are considered positives when $\geq 10\%$ of tumor cells show ER- and PR-specific staining by immunohistochemistry (IHC) ([Allison et al., 2020](#); [Dooijeweert et al., 2019](#)). For the **HER2-receptor status**, the HER2-staining by IHC is translated into a score depending on the stained cells percentage and staining intensity, and confirmed by *in situ* hybridization (ISH) when ambiguous ([Franchet et al., 2021](#); [Xu et al., 2019](#)). Hence, three main subgroups of breast carcinomas can be distinguished:

- **Hormone receptor-positive (HR⁺)** breast cancers express ER or PR or both and are thereby

STAGE	T - TUMOR	N - NODE	M - METASTASIS	DESCRIPTION
0	Tis	N0	M0	Carcinoma <i>in situ</i>
I	T1	N0	M0	Tumor < 2 cm
IIA	T1 T2	N1 N0	M0	Tumor < 2 cm with minor node invasion Tumor 2-5 cm without node invasion
IIB	T2 T3	N1 N0	M0	Tumor 2-5 cm with minor node invasion Tumor size > 5 cm without node invasion
IIIA	T1 T2 T3	N2 N2 N1/2	M0	Tumor < 2 cm with moderate node invasion Tumor 2-5 cm with moderate node invasion Tumor size > 5 cm with minor node invasion
IIIB	T4	N0/1/2	M0	Tumor of any size with direct extension to chest wall or skin with up to moderate node invasion
IIIC	Any T	N3	M0	Important node invasion
IV	Any T	Any N	M1	Distant metastases

TABLE 3 CLINICAL STAGES OF BREAST CANCER BASED ON THE TNM CLASSIFICATION.

sensitive to hormone therapy. The ER is expressed in 80% of breast cancers, and PR in 60-70% of breast cancers (Turashvili and Brogi, 2017).

- **HER2-positive** (HER2⁺) breast cancers present a gene amplification or overexpression of HER2 and are thereby sensitive to HER2 targeting therapies. They account for 15 to 20% of breast cancers.
- **Triple-negative** breast cancers (TNBC) are negative for these three markers, they are more aggressive than the other groups and insensitive to either hormone therapy or HER2-directed therapies. They account for 10 to 20% of breast cancers. Moreover, the TNBCs is a heterogeneous group that has been subdivided into six distinct subgroups: basal-like 1 (BL1), basal-like 2 (BL2), mesenchymal (M), mesenchymal stem-like (MSL), immunomodulatory (IM), and luminal androgen receptor (LAR) (Wang *et al.*, 2019a).

Other biomarkers can be used to further characterize breast carcinomas, including the androgen receptor (AR), the epidermal growth factor receptor (EGFR), p53, or Ki-67 (Silva *et al.*, 2019; Turashvili and Brogi, 2017).

MOLECULAR CLASSIFICATION

To better perceive the heterogeneity of breast cancers and improve tumor behavior prediction and associated therapies, breast carcinomas can also be classified by molecular patterns. Established by Perou *et al.* (2000), the molecular classification divides breast carcinomas into subtypes based on gene expression profiles similarities using the microarray technology. Hence, 4 subgroups have been defined by Perou *et al.* (2000): **luminal**, **HER2-enriched**, **normal-like**, and **basal-like**.

The luminal subtype, reflecting the expression pattern of a luminal cell of origin, was then divided into two subgroups: **luminal A** and **luminal B**, each having a distinct expression profile (Sørlie *et al.*, 2001). The existence of the normal-like subtype, characterized by a strong similarity with the normal mammary epithelial cells, is controversial as it is considered an artifact caused by the contamination with normal epithelial cells in the microarray analyses, thus it is now removed (Weigelt *et al.*, 2010b; Yersal and Barutca, 2014). Other subgroups have been defined over the years: the **claudin-low** subtype, introduced by Herschkowitz *et al.* (2007), is characterized by low expression of genes related to tight junctions, including claudins; and the **molecular apocrine** subtype, introduced by Farmer *et al.* (2005), is characterized by the expression of androgen receptor without expression of the ER and PR.

Summary of the different molecular subtypes (Makki, 2015; Weigelt *et al.*, 2010a; Yersal and Barutca, 2014):

- **Luminal A** subtype reflects the luminal origin of cancer and accounts for a majority of breast cancers (40-60%). Luminal A tumors express the luminal epithelial cytokeratins 8 and 18, ER and/or PR but not HER2. They have a low histological grade, relatively low proliferation, and a good prognosis.
- **Luminal B** subtype is also reflecting the luminal origin of cancer, and is characterized by a more aggressive phenotype, higher proliferation and higher histological grade (intermediate) than luminal A. Tumors are ER and/or PR -positive and approximately 30% of them are HER2-positive.
- **Molecular apocrine** is a rare subtype (less than 4% of all breast cancers) characterized by the overexpression of AR. Tumors are negatives for ER/PR and can be either HER2-positive or HER2-negative. Hence the HER2-negative apocrine tumors are triple-negative. They have an intermediate histological grade and a poor prognosis but a better one than other triple-negative breast carcinomas (Arciero *et al.*, 2020).
- **HER2-enriched** subtype is characterized by definition by a strong expression of HER2 and ER/PR is usually negative. HER2 positivity confers more aggressiveness, tumors are highly proliferative, have a high histological grade and a poor prognosis, but are sensitive to HER2-directed therapies.
- **Basal-like** subtype reflects the basal origin of cancer, as tumors express high levels of basal myoepithelial markers. Tumors are triple-negative for ER, PR, and HER2. They have a high histological grade, aggressive clinical behavior, and a high rate of metastasis.
- **Claudin-low** subtype is characterized by low expression of genes related to tight junctions and cell-cell adhesion, including claudins 3, 4, 7, occludin, and E-cadherin. Tumors are triple-negative for ER, PR, and HER2, and are highly expressing epithelial to mesenchymal transition genes and stemness related genes. They are highly proliferative, have a high histological grade, and have poor clinical outcomes.

Altogether, the different classifications are complementary and in constant evolution, a summary of the main breast cancer subgroups commonly referred to, with associated features, is rep-

resented in [Figure 20](#). The use of the different classifications, histological, clinical, and molecular is necessary to homogenize and improve therapeutic decision-making.

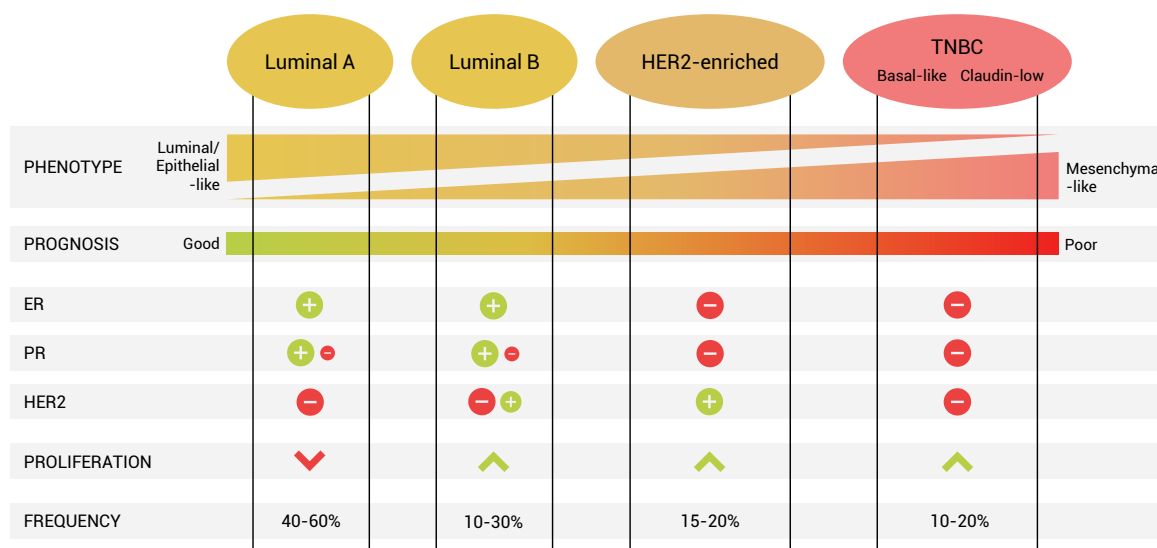


FIGURE 20 BREAST CANCER MAIN SUBGROUPS CHARACTERISTICS. The four main subgroups (or subtypes) of breast cancer along with their main characteristics: associated phenotype, prognosis, biomarkers expression (plus or minus), proliferation level (high or low), and estimated frequency ([Dai et al., 2015](#); [Eroles et al., 2012](#); [Kumar et al., 2015](#); [Niklaus et al., 2021](#); [Parise and Caggiano, 2014](#); [Weigelt et al., 2010a](#)).

3.2.5 THERAPIES

Different therapies are used to treat breast cancers: surgery, radiotherapy, chemotherapy, hormone therapy, and targeted therapies. The therapeutic choice depends on the cancer subtype, the degree of lymph nodes invasion, the expression of specific biomarkers such as the hormone receptors or HER2, the presence of metastases, and patient age and condition.

In the case of a non-metastatic breast cancer diagnosis, the **surgery** is systematically performed ([Figure 21](#)). The surgery can be partial (tumorectomy, removal of the tumor) or total (mastectomy, removal of the mammary gland) depending on the tumor size, and some adjacent lymph nodes can also be surgically removed.

A **chemotherapeutic treatment** can be delivered to reduce the tumor size prior to surgery (neoadjuvant therapy) or after surgery (adjuvant therapy) ([Leon-Ferre et al., 2021](#); [Spring et al., 2022](#)). Common chemotherapeutic drugs are anthracyclins (doxorubicin, epirubicin), cyclophosphamides, and taxanes (paclitaxel), generally used in combination (e.g. doxorubicin and cyclophosphamide followed by paclitaxel ([Esposito et al., 2014](#); [Lee and Nan, 2012](#); [Núñez et al., 2016](#)). Anthracyclins impair the DNA structure, preventing DNA replication and causing apoptosis, while taxanes suppress the microtubule formation, blocking the cell division and leading to apoptosis ([Conte et al., 2000](#); [Nabholtz and Gligorov, 2005](#)).

Coupled with the neoadjuvant or adjuvant chemotherapy, the **anti-HER2 therapy** is used

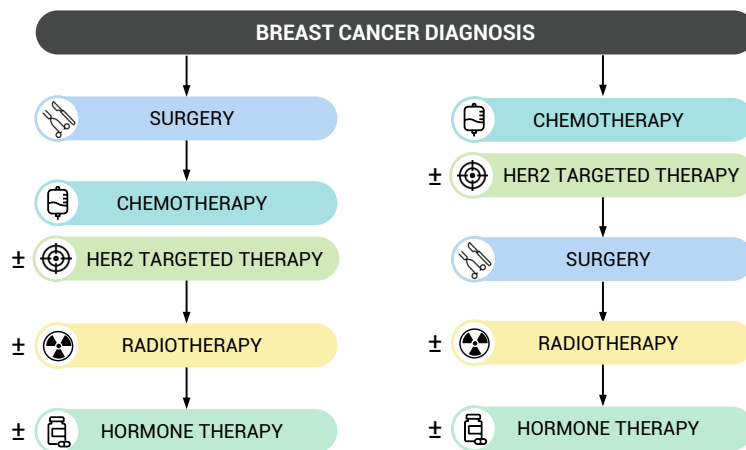


FIGURE 21 SCHEME OF CONVENTIONAL THERAPEUTIC CARE TO TREAT NON-METASTATIC BREAST CANCER.

against HER2-enriched breast cancers and is either composed of a monoclonal anti-HER2 antibody (trastuzumab, pertuzumab) or a tyrosine kinase inhibitor (lapatinib) (Escrivá-de-Romaní *et al.*, 2018; Wuerstlein and Harbeck, 2017).

The **radiotherapy** is a regional treatment essential to reduce the risk of relapse of breast cancers after surgery (Clarke *et al.*, 2006a). The radiotherapy treatment can be directly delivered during surgery for high precision irradiation of the targeted area, it is called intraoperative radiotherapy (Pilar *et al.*, 2017; Vaidya *et al.*, 2004). Then, after surgery, the conventional postoperative radiotherapy protocol is used in a fractionated deliverance of a total of 50 Gy, split into 25 fractions of 2 Gy delivered over 5 weeks (5 fractions per week) (Saksornchai *et al.*, 2021; Speers and Pierce, 2016). However, moderate hypofractionated and ultra-hypofractionated radiotherapy protocols have emerged due to the radiotherapy precision improvements allowing to deliver of higher doses to the tumor while limiting the dose received by surrounding normal tissues, and they are more convenient in terms of patient compliance and cost. Moderate hypofractionation uses fewer but higher dose fractions, commonly 40 Gy in 15 fractions over 3 weeks (e.g. the START trial, Bentzen *et al.*, 2008a; Bentzen *et al.*, 2008b; Haviland *et al.*, 2013) while ultra-hypofractionation uses even higher doses per fraction, commonly 26 Gy in 5 fractions over 1 week (e.g. the FAST-Forward trial, Brunt *et al.*, 2016; Brunt *et al.*, 2020; Brunt *et al.*, 2021) (Kim *et al.*, 2021; Saksornchai *et al.*, 2021; Postoperative radiotherapy for breast cancer: hypofractionation RCR consensus statements, 2021).

The **hormone therapy** is used to treat hormone-dependent tumors such as the luminal A and B subtypes expressing the ER and PR. Selective estrogen receptor modulators (SERM) are ER ligands acting like estrogens in some tissues but blocking their action in others. The SERM-induced conformational changes of the estrogen receptor (ER) change its interactions with coactivators or corepressors and subsequently initiate or suppress the transcription of the ER target genes (Lewis and Jordan, 2005; Shelly *et al.*, 2008). Tamoxifen is a SERM that has been clinically used for the last 40 years to treat ER-positive breast cancers, but the acquisition of tamoxifen resistance (tumors

becoming hormone-independent) remains a major challenge to improving breast cancer treatment efficacy (Ali *et al.*, 2016; Chang, 2012).

3.3 BREAST CANCER STEM CELLS

3.3.1 MARKERS AND ISOLATION

The study of CSCs requires markers to identify, quantify, and isolate them. Markers used for breast CSCs are similar to those used for normal stem cells but are not universal either between tumors or within the same tumor. Thereby, depending on the cell line or model studied, some markers are more fitted than others. This CSC phenotype heterogeneity thus requires the use of several markers completed with functional validation of their stem cell capacities, such as their self-renewal ability.

SURFACE MARKERS

The expression of surface markers is one of the first means used to identify CSCs (Bonnet and Dick, 1997). In breast cancers, several markers can be used to isolate the CSC population. The most common combination of markers used is a high expression of CD44 coupled with a low expression of CD24. The CD44^{high}/CD24^{-low} population was one of the first to be found with stem cell characteristics (Al-Hajj *et al.*, 2003).

CD44 is a membrane glycoprotein and functions as a hyaluronic acid receptor involved in the regulation of cell migration, adhesion, proliferation, and survival while CD24 is a surface protein involved in cell adhesion (Orlan-Rousseau, 2015).

The CD44^{high}/CD24^{-low} labeling identifies a tumorigenic cell population capable of regenerating a heterogeneous population, by forming mammospheres *in vitro* or tumors *in vivo* (Al-Hajj *et al.*, 2003; Ponti *et al.*, 2006). However, this combination of markers is not usable for all breast cancer subtypes and lineages, and the proportions of CD44^{high}/CD24^{-low} cells vary between different cell lines. Indeed, these markers are adequate for the luminal subtypes, but not for the basal one as the majority of cells are CD44^{high}/CD24^{-low} (nearly 100% for the MDA-MB-231 cell line) (Fillmore and Kuperwasser, 2008; Stuelten *et al.*, 2010). Yet, this labeling is still widely used in the literature.

CD44 and CD24 are also used as prognosis biomarkers for breast cancers. The CD44⁺/CD24⁻ phenotype predicts a poor clinical outcome in TNBC but can be considered a favorable prognostic in hormone receptor-positive breast cancers (Ahmed *et al.*, 2012; Giatromanolaki *et al.*, 2011; Kim *et al.*, 2011; Wang *et al.*, 2017a).

Additionally, the CD133 transmembrane protein is a CSC marker in several cancers (e.g. in glioblastoma), and used in combination with other markers in breast cancers for CSC population identification (Crocker *et al.*, 2008; Joseph *et al.*, 2019; Mansour and Atwa, 2015).

ENZYMATIC ACTIVITY

The aldehyde dehydrogenase (ALDH) is an enzyme responsible for the oxidation of aldehydes into carboxylic acids and is involved in the oxidative metabolism of retinol into retinoic acid. The ALDH activity is higher in MaSCs and mammary CSCs and is associated with poor prognosis and resistance to treatment (Croker *et al.*, 2008; Ginestier *et al.*, 2007; Tanei *et al.*, 2009). Hence, the Aldefluor test has been developed to identify normal and cancerous stem cells based on the ALDH enzymatic activity detection (Ginestier *et al.*, 2007). This test relies on a non-fluorescent substrate, which becomes fluorescent when cleaved by ALDH enzymes.

ALDH^{+/high} cells are characterized by increased tumorigenicity as well as self-renewal and differentiation capabilities. In addition, the ALDH inhibition sensitizes ALDH^{high}/CD44⁺ CSCs to chemotherapy and radiation therapy (Croker and Allan, 2011). In breast CSCs, the high ALDH activity is mainly due to the ALDH1A1 and ALDH1A3 isoforms (Marcato *et al.*, 2011a; Marcato *et al.*, 2011b). The expression of a fluorescent protein under the control of the ALDH1A1 promoter has been used as a CSC reporter system, allowing the live tracking of breast CSCs (Bidan *et al.*, 2019). The ALDH activity is also used to detect CSCs in other cancer types such as bladder and cervical cancers (Xu *et al.*, 2015).

SIDE POPULATION

The excluded population, or “side population”, is based on the high activity of efflux pumps of the ABC transporter family, which are over-represented at the membrane of normal and cancer stem cells (Britton *et al.*, 2012). It can be identified by flow cytometry using a fluorescent vital dye exclusion test such as Hoechst 33342 or Rhodamine 123. These membrane transporters are notably involved in the efflux of chemotherapeutic drugs by CSCs and are thus associated with chemoresistance (Wu and Alman, 2008; Zhou *et al.*, 2001). However, the side population consists of a CSC-enriched population, and not all CSCs can be identified by the ability to exclude vital dyes (Behbod and Vivanco, 2015; Britton *et al.*, 2012; Golebiewska *et al.*, 2011). This demonstrates the requirement for marker combinations to isolate CSCs.

LOW PROTEASOME ACTIVITY

Several cellular functions are regulated by the proteasome, such as DNA repair mechanisms and cell cycle regulations. CSCs have been shown to have low proteasome activity in several cancer models, including breast cancer (Vlashi *et al.*, 2009). Hence, the use of a proteasome activity reporter allows for the detection and isolation of a CSC-enriched population (Vlashi *et al.*, 2013; Vlashi *et al.*, 2014). A reporter system used in breast cancer consists in the stable expression of a fusion fluorescent protein, composed of the C-terminal degron of murine ornithine decarboxylase (cODC) and a fluorescent protein ZsGreen, which is specifically degraded by the proteasome, thus, an accumulation of fluorescence reveals a low proteasomal activity (Hoyt *et al.*, 2005).

3.3.2 FUNCTIONAL TESTS

In addition to the expression of specific markers, functional tests have also been developed and are recognized as essential tests to define a stem cell population: the **SFC** *in vitro* test and the **tumorigenicity** *in vivo* test.

SPHERE FORMING CAPACITY IN VITRO ASSAY

The sphere-forming capacity (SFC) test is based on the ability of stem cells to grow in the absence of anchorage. The cells are seeded in small quantities in a specific sphere-forming medium, enriched in growth factors and without serum, and under non-adherent culture conditions. The number of spheres formed is quantified to estimate the percentage of sphere formation (Dontu *et al.*, 2003; Lombardo *et al.*, 2015). The sphere culture allows for the selection of both CSCs and progenitors in the first step. Then, spheres can be cultivated over several generations to eliminate progenitor cells, as the sphere-forming CSCs have the ability to self-renew over generations.

TUMORIGENICITY IN VIVO ASSAY

The tumorigenicity test is based on the self-renewal, differentiation, and tumor-initiating capacities of CSCs. A prospective cell population is injected into mice in limited numbers of cells. The presence or absence of tumor development defines the stemness of the tested population, or at least its enrichment in CSCs, as only CSCs should be able to regenerate a tumor. Important features observed are the abilities of these cells to form tumors in series, revealing their self-renewal capacity, and to reproduce the heterogeneity of the initial tumor, revealing their differentiation capacity (O'Brien *et al.*, 2010). This technique remains the best method for characterizing tumor-initiating cells and also allows for the study of cell differentiation capacity by examining the heterogeneity of newly formed tumors. (Visvader and Lindeman, 2008).

KEY POINTS

- Above 2 millions of new breast cancers cases are diagnoses each year. It is the most diagnosed cancer worldwide, and the first cause of cancer mortality for women.
- Breast tumors are heterogeneous, hence several classifications relying on different parameters, histological, clinical and molecular ones, have been established to improve clinical outcome predictions and therapeutic decisions.
- Breast cancers can be divided into four main subtypes based on their molecular patterns: luminal A, luminal B, HER2-enriched and TNBC, or defined by their biomarker expression profiles as ER/PR-positive, HER2-positive, and triple-negative.
- Breast CSCs can be identified by several markers such as surface markers (CD44^{high}/CD24^{-/low}), enzymatic activity (ALDH^{high}), exclusion of vital dye by efflux pumps (Hoechst side population), or low proteasome activity. In addition, to confirm their stem cell properties, functional tests must be performed such as the sphere-forming capacity (SFC) *in vitro* assay and the tumorigenicity *in vivo* assay.

[^ Back to Table of Contents](#)

4

DNA METHYLATION: PRINCIPLES AND METHODOLOGY

☰ CHAPTER CONTENTS

4.1	DNA METHYLATION	105
4.1.1	PRINCIPLE	105
4.1.2	ESTABLISHMENT, MAINTENANCE, AND REMOVAL	106
4.1.3	ROLE OF DNA METHYLATION	108
4.1.4	DNA METHYLATION AND EPIGENETIC REPROGRAMMING	109
4.1.5	DNA METHYLATION IN CANCER	110
4.1.6	DNA METHYLATION AND STEMNESS	111
4.1.7	EFFECTS OF RADIATIONS ON DNA METHYLATION	112
📌	KEY POINTS	114
4.2	OVERVIEW OF METHODS TO STUDY DNA METHYLATION	114
4.2.1	BISULFITE-BASED ASSAYS	115
4.2.2	ENZYME DIGESTION-BASED ASSAYS	121
4.2.3	AFFINITY ENRICHMENT-BASED ASSAYS	122
📌	KEY POINTS	123
4.3	BISULFITE SEQUENCING PCR (BSP)	123
4.4.1	EXPERIMENTAL PROCESS	124
📌	KEY POINTS	130
4.4.2	ANALYSIS STRATEGY AND TOOLS	130
📌	KEY POINTS	139
4.4.4	ARTIFACTS	139
📌	KEY POINTS	147

[^ Back to Table of Contents](#)

4.1 DNA METHYLATION

4.1.1 PRINCIPLE

DNA methylation is characterized by the presence of an additional **methyl group** at the **5' position of cytosine residues** (Figure 22). The 5-methylcytosine (5mC) is a stable and heritable alteration that does not involve modifications of the DNA sequence itself, and has a regulatory role

on gene expression, thereby it is classified as an epigenetic mark.

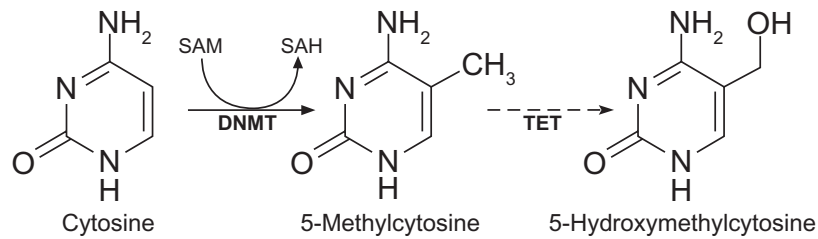


FIGURE 22 METHYLATION OF CYTOSINE IN DNA. The DNA methyltransferases (DNMTs) convert cytosine to 5-methylcytosine by adding a methyl group, donated from S-adenosylmethionine (SAM) molecules. The oxidation of 5-methylcytosine by the TET family of enzymes generate 5-hydroxymethylcytosine. SAH = S-adenosylhomocysteine. Figure from [Greer and McCombe \(2012\)](#).

In vertebrates, the cytosine methylation mainly occurs at CG dinucleotides, cytosine followed by guanine, called **CpG** (cytosine-phosphate-guanine) sites. But, non-CpG methylation has also been found in stem cells ([Lister *et al.*, 2009](#); [Ziller *et al.*, 2011](#)). In plants, the cytosine methylation occurs in several contexts: CG, CHG (C followed by a non-G followed by a G), and CHH (C followed by two non-Gs) ([Chan *et al.*, 2005](#); [Lister *et al.*, 2008](#); [Zhang *et al.*, 2006](#)).

The human genome contains ~29 millions of cytosine-phosphate-guanines (CpGs) and 60% to 80% of them are methylated ([Kim and Costello, 2017](#); [Lister *et al.*, 2009](#); [Tost, 2010](#)). In total, the 5-methylcytosines represent ~1% of all bases of the human genome ([Ehrlich *et al.*, 1982](#); [Tost, 2010](#)).

The CpG sites can be clustered into regions named **CpG islands (CGIs)**, which are regions of high CG density. Approximately 7% of CpGs are located in CGIs ([Deaton and Bird, 2011](#)). Several definitions of a CGI exist and differ due to a few variations of parameters, but in general, a CpG island corresponds to a 200 or 500 bp minimum sized region with a GC content higher than 50% and an observed *versus* expected ratio higher than 0.6 for the occurrence of CpG sites ([Gardiner-Garden and Frommer, 1987](#)). These islands extent is generally from 0.5 to 3 kb, and are estimated to be ~30,000 in the human genome, accounting for ~2% of the genome ([Cross and Bird, 1995](#); [Tost, 2010](#)).

4.1.2 ESTABLISHMENT, MAINTENANCE, AND REMOVAL

For DNA methylation patterns to be established, maintained or erased, the mammalian **DNA methyltransferase (DNMT)** family of enzymes catalyze the transfer of a methyl group from S-adenosylmethionine (SAM) to cytosine, while the **ten-eleven translocation (TET)** family proteins catalyze the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) ([Figure 22](#)) ([Chen and Riggs, 2011](#); [Ginno *et al.*, 2020](#); [Jones and Liang, 2009](#); [Meng *et al.*, 2015](#)).

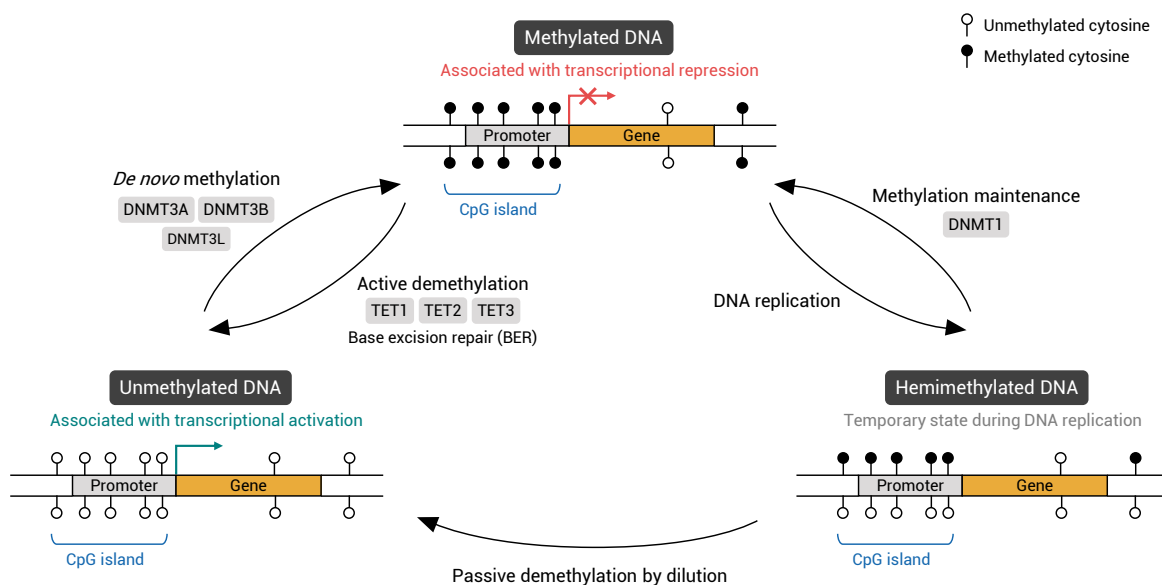


FIGURE 23 REPRESENTATION OF DNA METHYLATION STATES WITH ENZYMES INVOLVED.

MAINTENANCE OF DNA METHYLATION

DNA methylation motifs are transmitted through DNA replication with high fidelity. During replication, the template strand is still methylated while the newly synthesized strand lacks 5-methylcytosines, the DNA is hemimethylated (Figure 23). Among the DNMT family of enzymes, the DNMT1 restores the symmetrical DNA methylation, as it has a strong preference for hemimethylated CpG sites, by copying the pre-existing methylation patterns to the new strand (Chen and Li, 2004; Chen and Riggs, 2011; Pradhan *et al.*, 1999). The DNMT1-mediated DNA methylation maintenance involves the ubiquitin-like protein with PHD and ring finger domains 1 (UHRF1) for recognition of hemimethylated DNA. Indeed, by binding to hemimethylated sites, the hemimethylated DNA-binding protein UHRF1 helps recruiting DNMT1 to DNA replication foci to methylate cytosines of the newly synthesized strand (Bostick *et al.*, 2007; Bronner *et al.*, 2019).

DE NOVO DNA METHYLATION

To establish new methylation patterns, *de novo* methylation on unmethylated CpG sites is required. The DNMT3A and DNMT3B are methyltransferases active on unmethylated DNA and responsible for the establishment of methylation motifs during early development (Figure 23) (Edwards *et al.*, 2017). These two enzymes have no preference for hemimethylated DNA, though they can also participate in the maintenance of DNA methylation (Chen and Riggs, 2011; Okano *et al.*, 1999).

The activity of DNMT3A and DNMT3B can be enhanced by the DNMT3L protein, which does not have a methyltransferase activity itself, by direct interaction and complex formation to stimulate *de novo* DNA methylation (Chen *et al.*, 2005; Jia *et al.*, 2007; Suetake *et al.*, 2004).

The establishment of new DNA methylation is strongly linked to histones post-translational

modifications epigenetic marks. The DNMT3A and DNMT3B can be recruited by proteins associated with histone marks such as G9A, the histone methyltransferase responsible for histone H3 lysine 9 (H3K9) mono- and dimethylation (Epsztejn-Litman *et al.*, 2008). And, for example, *de novo* methylation is associated with the PRC2 (polycomb repressive complex 2) activity (Viré *et al.*, 2005). PRC2 is responsible for the H3K27me₃, a repressive mark of transcriptional activity. Also, *de novo* methylation is associated with the histone deacetylase (HDAC) activity and the lysine demethylase (KDM)1 activity which induces the histone H3 lysine 4 (H3K4) demethylation (Dobosy and Selker, 2001; Wang *et al.*, 2009a). Both H3K4me₃ and histone acetylation (HAc) are marks associated with transcriptional activation, therefore their removal is associated with transcriptional repression.

REMOVAL OF DNA METHYLATION

The erasure of DNA methylation patterns can be passive, through the absence of DNA methylation maintenance during DNA replication, or active through enzymatic reaction mediated by TET enzymes (Figure 23) (Chen and Riggs, 2011). The TET proteins, composed of TET1, TET2 and TET3 proteins, can oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC).

Then, the 5-hydroxymethylcytosine (5hmC) can be converted to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TET enzymes. 5-carboxylcytosine (5caC) is then considered as a defective base and is excised by base excision repair (BER) pathways *via* the thymine-DNA glycosylase (TDG), completing the DNA demethylation process (He *et al.*, 2011; Ito *et al.*, 2011; Shen *et al.*, 2013).

4.1.3 ROLE OF DNA METHYLATION

Although approximately 7% of CpG sites are located in CGIs, these are prevalent in gene promoters (Deaton and Bird, 2011). Approximately 70% of the gene promoters have a high CpG concentrations (Saxonov *et al.*, 2006). And, their hypermethylation is frequently associated with transcriptional repression (Figure 23) (Lande-Diner *et al.*, 2007). Moreover, the distal regulatory elements such as enhancers or silencers, capable of binding transcription regulation factors, are also subjected to DNA methylation; thereby, intergenic DNA methylation also contributes to the gene expression regulation (Li *et al.*, 2018a; Schübeler, 2015).

Methylated cytosines can interact with specific proteins, triggering alterations of the chromatin structure, affecting the transcription rate of the genes nearby (Jones and Takai, 2001; Moore *et al.*, 2013; Schübeler, 2015). Proteins containing an methyl-CpG-binding domain (MBD), such as MBD1, MBD2, and MeCP2, can interact with methylated cytosines and recruit proteins, transcription repression factors, or chromatin modifiers including HDAC, all driving to repression of the transcription (Baubec *et al.*, 2013; Klose and Bird, 2006; Nan *et al.*, 1998).

Jones and Takai (2001) explain that the methylation of cytosines can serve as an efficient and heritable silencing mechanism without relying on DNA-protein interactions. Indeed, the DNA

methylation profile conveys information on gene regulation from parent cell to daughter cells through cellular replication, without alteration of the DNA sequence. It provides an additional layer of information enabling the establishment of epigenetic marks, according to [Mitchell *et al.* \(2016\)](#). In this way, the DNA methylation maintains gene regulatory programs and function as a cellular memory ([Kim and Costello, 2017](#); [Shipony *et al.*, 2014](#)). The DNA methylation is notably required for the expression regulation of imprinted genes — genes that show differential expression between the parental alleles in diploid cells — as well as for the X chromosome inactivation ([Barlow and Bartolomei, 2014](#); [SanMiguel and Bartolomei, 2018](#)).

The DNA methylation, along with the cellular epigenome, safeguards the cellular differentiation, identity, and diversity, critical features required for a normal development, where pluripotent cells differentiate into a variety of cell types ([Kim and Costello, 2017](#); [Okano *et al.*, 1999](#)).

4.1.4 DNA METHYLATION AND EPIGENETIC REPROGRAMMING

The study of embryonic stem cells (ESCs) provided many insights on the role of DNA methylation in early development and cell differentiation. A wave of global resetting of DNA methylation occurs in the early embryo and determines the establishment of the methylation patterns ([Edwards *et al.*, 2017](#); [Lienert *et al.*, 2011](#); [Smith and Meissner, 2013](#)). The CGIs located in promoters of many genes necessary for development are found hypomethylated in ESCs, while the CGIs in intergenic regions are frequently subjected to DNA methylation during development, suggesting potential regulatory functions ([Illingworth *et al.*, 2010](#)). The hypomethylation of promoter-associated CGIs suggests the exclusion of DNA methyltransferases and transcription factor binding to initiate transcription of genes required for development ([Brandeis *et al.*, 1994](#)).

Several studies have shown the presence of specific regions or elements, such as transcription factors binding sites, contributing to DNA methylation states, particularly for the maintenance of unmethylated CGIs at promoters in ESCs, and for their differentiation ([Brandeis *et al.*, 1994](#); [Dickson *et al.*, 2010](#); [Lienert *et al.*, 2011](#)). Genes required for the pluripotency of ESCs, such as NANOG and OCT4 (POU5F1), switch from an unmethylated state associated with active transcription to a methylated state, associated with their transcriptional repression upon differentiation ([Deb-Rinker *et al.*, 2005](#); [Yeo *et al.*, 2007](#)). Additionally, DNA methylation and histones modifications have been shown to inter-operate to establish specific chromatin conformations, through context-dependent cross-talk and addition of mutually exclusive marks or co-regulatory marks ([Cedar and Bergman, 2009](#)). These epigenetic marks determine the lineage commitment of cells during development.

Once the DNA methylation patterns and the chromatin structures modulating the DNA accessibility are established during development, the cell-type-specific gene expression programs are stabilized. The expression of stem-specific or lineage-unrelated genes is prevented by the epigenome, thus preserving the cell phenotype and its inheritance ([Mohn *et al.*, 2008](#)). However, the epigenetic reprogramming through global DNA demethylation — through single-stranded DNA breaks and

BER — is also required for the germ line formation (primordial germ cells, PGCs), to erase the genomic imprint and to return cells to a pluripotent state (Reik, 2007). Thereby, as explained above, there are two waves of epigenetic reprogramming in the mammalian life cycle, one during the early embryogenesis, establishing the expression patterns in cells for proper development of the organism, and the other for the germ line formation, resetting the epigenome in preparation for the future offspring (Xavier *et al.*, 2019; Zeng and Chen, 2019).

Artificially induced reprogramming of somatic cells into stem-like cells has been possible by the reexpression of four transcription factors — OCT4, SOX2, C-MYC, and KLF4 — giving rise to iPS cells (Takahashi *et al.*, 2007; Takahashi and Yamanaka, 2006). Along with the acquisition of stem cell properties, global DNA methylation profiles changes have been observed; the profiles are similar to those in ESCs, despite residual methylation signatures from the original donor state of the iPS cell (Kim *et al.*, 2010; Nishino and Umezawa, 2016). While most of the DNA methylation changes are towards the hypermethylation of CpG sites in iPS cells, for a limited number of genes associated with stemness, such as OCT4 and NANOG, their promoters lose their DNA methylation marks (Olariu *et al.*, 2016; Takahashi *et al.*, 2007; Takahashi and Yamanaka, 2006).

4.1.5 DNA METHYLATION IN CANCER

The cancer cells and normal stem cells present DNA methylation differences. In general, a decrease of CpG methylation is observed in tumor cells, but both hypo- and hypermethylation events have been reported. Indeed, tumor suppressor genes are silenced by DNA methylation during tumorigenesis while proto-oncogene are upregulated by removal of DNA methylation (Baylin and Jones, 2016; Kulis and Esteller, 2010). For example the promoter of the tumor suppressor gene BRCA1 gene is generally hypermethylated in breast cancers (Esteller *et al.*, 2000). Hanahan (2022) integrated in the hallmarks of cancer the “nonmutational epigenetic reprogramming” hallmark, defined as a genome reprogramming exclusively caused by epigenetically regulated changes in gene expression, meaning that mutations are not always required for cancer development and progression (Darwiche, 2020; Huang, 2012; Nam *et al.*, 2021). For example, hypoxia-induced TET-mediated epigenetic changes could drive the initiation of tumorigenesis if the cell-of-origin resides in a hypoxic area (Hanahan, 2022; Michealraj *et al.*, 2020; Thienpont *et al.*, 2016). Hence, DNA methylation changes participates in cancer initiation and progression.

The DNA methylation patterns of cancer cells can serve as biomarkers for diagnosis. Indeed, the measurements of DNA methylation associated to genes, such as bone morphogenetic protein (BMP) 3, TWIST1, or O₆-methylguanine-DNA methyltransferase (MGMT) coding for a repair protein, are used in biomarker assays (Caccese *et al.*, 2022; Kessel *et al.*, 2016; Koch *et al.*, 2018).

To force the reexpression of genes with an hypermethylated promoter in cancer cells, especially genes required for the cancer cell survival, demethylation agents have been used as therapeutic agents in AML (Dombret *et al.*, 2015; He *et al.*, 2014; Wongtrakoongate, 2015). The 5-azacytidine

(also known as 5-aza-CR, azacitidine or Vidaza) and 5-aza-2'-deoxycytidine (also known as 5-aza-CdR, decitabine or Dacogen) agents are two nucleoside analogs of the cytosine base. They are incorporated during DNA replication and, as the binding of DNMTs to these analogs by covalent interaction is irreversible, it triggers the DNMTs excision from DNA and its subsequent degradation by the proteasome, resulting in DNMT depletion and methylation inhibition after several rounds of DNA synthesis (Ghoshal *et al.*, 2005; Mehdipour *et al.*, 2020).

4.1.6 DNA METHYLATION AND STEMNESS

DNA methylation regulates the self-renewal and pluripotency of both normal and cancer stem cells. First, in normal cells, NANOG and OCT4 genes are regulated by DNA methylation throughout the differentiation (Fouse *et al.*, 2008; Hattori *et al.*, 2007; Li *et al.*, 2007b).

Indeed, Deb-Rinker *et al.* (2005) have reported an increase in DNA methylation at the NANOG upstream region during the retinoic acid-induced differentiation of NT2 cells, which is correlated with a decrease in NANOG expression (pluripotent cell line) into neurons. Similarly, Li *et al.* (2007b) observed an hypermethylation of OCT4 and NANOG promoter regions upon retinoic-acid-induced differentiation of ESC. At day 0, the OCT4 promoter is methylated at 0% and the NANOG promoter at 7.6%, however, after 3 days of induced differentiation, the OCT4 promoter methylation increased up to 22.4% and the NANOG promoter up to 59.7%. Additionally they highlighted the involvement of de novo methylation by DNMT3A and DNMT3B as the knock-down of these two genes abolished the differentiation-induced hypermethylation of both promoter regions of OCT4 and NANOG. Indeed, after 3 days, the OCT4 and NANOG promoter methylation levels are 4.7% and 17.2% in DNMT3A knockdown cells, 3.6% and 10.1% in DNMT3B knockdown cells, and 0% and 3% in the double-knockdown cells, respectively (Li *et al.*, 2007b). During ESC differentiation, the orphan nuclear receptor germ cell nuclear factor (GCNF) binds to the OCT4 upstream region and recruits MBD proteins and DNMT3A, resulting in hypermethylation and repression of OCT4 (Gu *et al.*, 2005; Gu *et al.*, 2011; Wang *et al.*, 2016). Furthermore, the DNA methylation status of NANOG and OCT4 promoters have been correlated with their expression levels throughout the differentiation of ESC (Hattori *et al.*, 2007; Wang *et al.*, 2009b). In addition, the epigenetic regulation of OCT4 and NANOG also involve histones modifications and chromatin remodeling (Fouse *et al.*, 2008; Hattori *et al.*, 2007; Kashyap *et al.*, 2009; Topalovic *et al.*, 2017). In cancer stem cells, the promoter region of NANOG have also been found hypomethylated (Liu *et al.*, 2020c; Wang *et al.*, 2013c). Indeed, Wang *et al.* (2013c) observed lower methylation levels of the NANOG promoter in CSCs compared to non-CSCs (42% and 89% respectively), correlated to an upregulation of its expression in hepatocellular carcinoma.

Aside from pluripotency factors promoters, stem cells exhibit a different methylation profile than differentiated cells, both in normal and cancer cells (Bibikova *et al.*, 2006; Bock *et al.*, 2012; Helou *et al.*, 2014; Lee *et al.*, 2015; Meissner *et al.*, 2008; Yu *et al.*, 2020). Bock *et al.* (2012)

describe DNA methylation patterns as a fingerprint of the cell phenotype which includes the pluripotent state. The CSC-specific methylation patterns have been studied in several models, including esophageal cancer, glioblastoma and breast cancer, in which several DMRs between non-CSCs and CSCs are identified (Helou *et al.*, 2014; Lee *et al.*, 2015; Li *et al.*, 2018b; Yu *et al.*, 2020). By a whole-genome bisulfite sequencing (WGBS) comparative analysis between breast CSCs and non-CSCs (MDA-MB-231 cell line), Li *et al.* (2018b) identified 8,007 genes associated with an hypermethylated promoter and 6,175 genes associated with an hypomethylated promoter. The gene ontology enrichment analysis revealed association with cell development processes and cell–cell signaling. Additionally, the integrative co-analysis of the transcriptome, methylome and histone modifications highlighted 23 genes, including 8 downregulated tumor suppressor genes, with a consistent regulation from both promoter DNA methylation and histone modification at transcription start site (TSS) correlated with transcriptomic expression. The PTPN14, WWC1, NOS1AP, OBSCN, FAM189B, FHL3, ROBO1, ARHGAP10, ARAP1, FRS3, NAV2, MAPK13, MTMR3, RERE, and AC093838.4 genes are found hypermethylated and downregulated, and the PKN2, CRCP, TMEM71, PPTC7, R3HDM1, CDC42EP3, SLC6A3, and PTPRN2 are found hypomethylated and upregulated in CSCs (Li *et al.*, 2018b).

4.1.7 EFFECTS OF RADIATIONS ON DNA METHYLATION

The radiation exposure, either a single dose or fractionated doses, induces DNA methylation changes in cancer cells (Antwih *et al.*, 2013; Bae *et al.*, 2015; Danielsson *et al.*, 2020; Kuhmann *et al.*, 2011; Miousse *et al.*, 2017). For instance, Antwih *et al.* (2013) irradiated MDA-MB-231 breast cancer cells at 2 or 6 Gy and analyzed DNA methylation differences from 1 to 72 hours post-irradiation, using the Illumina Infinium 450K methylation profiling array. They observed differentially methylated genes associated with cell cycle regulation, DNA repair mechanisms and apoptosis pathways, indicating the involvement of DNA methylation in the cellular response to irradiation. Notably, the IGF1R (tyrosine kinase receptor), KRAS (GTPase, oncogene) and HDAC4 (histone deacetylase) genes were found differentially methylated after 6 Gy irradiation. Another study, Kuhmann *et al.* (2011), used a fractionated radiation treatment on MCF7 breast cancer cells (5×2 Gy per week, for a total of 10 Gy and 20 Gy) and identified differentially methylated genes by methyl-CpG immunoprecipitation followed by CGI microarray. From this analysis, they selected 15 CGIs differentially methylated for further validation by MassARRAY, and confirmed a methylation increase at the ADAMTS9 (metallopeptidase) promoter, FOXC1 (transcription factor) gene, and intragenic CGI in the TRAPPC9 (trafficking protein), and a methylation decrease at the AMIGO3 (adhesion molecule) promoter.

Moreover, the irradiation affects gene expression, that may include DNMTs in specific model and context (Miousse *et al.*, 2017). Mice exposed to radiations showed decreased expression of DNMTs associated to a general decrease of DNA methylation in hematopoietic tissues (Miousse *et al.*, 2014; Pogribny *et al.*, 2005). In the MDA-MB-231 breast cancer cell line, Antwih *et al.* (2013) reported

a gradual decrease of the DNMT1 protein levels in the 72h after 6 Gy irradiation. In nasopharyngeal carcinoma, [Wu et al. \(2020\)](#) found DNMT3B expression upregulated after exposure to radiations, mRNA and protein levels increasing gradually within the 48h following irradiation, while DNMT1 and DNMTA expressions were not consistently affected. They highlighted the role of DNMT3B in the radioresistance of these cells through methylation of genes coding for p53 and p21.

Furthermore, the radiation-induced DNA methylation changes have been associated to DNA repair mechanisms and genomic instability ([Antwih et al., 2013](#); [Armstrong et al., 2012](#); [Kaup et al., 2006](#); [Sutton et al., 2019](#)). [Sutton et al. \(2019\)](#) observed that prostate cancer cells with different responses to DNA damage have also distinctive DNA methylomes. Indeed, the resistant cells generally exhibited higher levels of DNA methylation, but it does not influence their susceptibility to DNA damage or the rate of short-term repair. [Armstrong et al. \(2012\)](#) found that global methylation levels do not determine the radiosensitivity of mouse ESCs. However they reported the involvement of DNMTs in the radiation-induced genomic instability, especially DNMT1, which is recruited to DNA repair sites, as its knock-down increased the mutation rate, potentially due to the absence of methylation at specific regions, hindering the DNA repair mechanisms ([Armstrong et al., 2012](#)).

Finally, DNMT inhibitors have been used in combination with radiotherapy to radiosensitize tumors by potentiating the radiations cytotoxic effects ([Gravina et al., 2010](#)). As the cytosine analogs forms covalently bound DNA-protein complexes with DNMTs, it may be more difficult to repair proximal DNA damage, thus increasing the cytotoxicity ([Ferguson et al., 1997](#); [Jüttermann et al., 1994](#); [Kim et al., 2012](#)). In addition, DNMT inhibitors have been shown to synchronize cancer cells preferentially in the G1 or G2/M phase of the cell cycle, which are the most radiosensitive phases ([Qiu et al., 2009](#)). Plus, the DNMT inhibitors modulates gene expression and signaling pathways, such as NF- κ B signaling and apoptosis signaling, that also potentiate the cytotoxic effect of radiotherapy ([Khong et al., 2008](#); [Peitzsch et al., 2016](#); [Zhu et al., 2018](#)).

KEY POINTS

- The DNA methylation consists of the addition of a methyl group to cytosines, located at CpG sites (CG dinucleotides) in vertebrates.
- The methylation and demethylation dynamics are represented in [Figure 23](#). The DNA methyltransferase (DNMT) family of enzymes are responsible for DNA methylation and the ten-eleven translocation (TET) family of enzymes for demethylation.
- Regions rich in CpG are called CpG island (CGI) and are preferentially found at promoter loci. Their hypermethylation is generally associated with transcriptional repression.
- As DNA methylation marks are heritable through cellular divisions, it serves as an epigenetic memory, transmitting the regulation of gene expression programs without alteration of the DNA sequence, critical to maintaining the cell identity.
- Epigenetic reprogramming events, involving demethylation and remethylation of DNA, occur in the early embryo and during the germ line formation, to establish the DNA methylation patterns regulating gene expression programs necessary for the development.
- In cancer, aberrant DNA methylation patterns can be found, such as hypermethylation of tumor suppressor genes and hypomethylation of proto-oncogenes.
- The stem state identity of cells involves a DNA methylation regulation in both normal and cancer cells. Indeed, in addition to the pluripotency factors expression epigenetic regulation, DNA methylation signatures of stemness have been identified in normal and cancer cells.

^ [Back to Table of Contents](#)

4.2 OVERVIEW OF METHODS TO STUDY DNA METHYLATION

A plethora of methods to study DNA methylation have been developed over the years and can be classified either by ① their principle (the way to discriminate 5-methylcytosine (5mC): by bisulfite conversion, by restriction enzyme digestion, by affinity binding, or by a combination of them), ② their methodology (experimental techniques used: polymerase chain reaction (PCR), Sanger sequencing, microarray, next-generation sequencing (NGS)...), ③ their resolution (global methylation levels to single CpG sites) or ④ their coverage (region-specific, array or wide profiling). A non-exhaustive classified list is presented in [Table 4](#).

Here, a classification of DNA methylation study techniques is proposed, divided in three major groups: ① bisulfite-based assays, ② restriction enzyme digestion-based assays and ③ affinity enrichment-based assays ([Khodadadi et al., 2021](#); [Laird, 2010](#); [Rauluseviciute et al., 2019](#)) ([Table 4](#)). Following this classification, the most known ones are described briefly below ([Chatterjee et al., 2017](#); [Khodadadi et al., 2021](#); [Li and Tollefsbol, 2020](#); [Rauluseviciute et al., 2019](#); [Soozangar et al., 2018](#)).

4.2.1 BISULFITE-BASED ASSAYS

The sodium bisulfite treatment of DNA converts unmethylated cytosines into uracils, while methylcytosines are not affected. During DNA amplification, uracils are replaced by thymines, and thereby, the cytosine methylation status can be revealed by the presence of either a cytosine (methylated) or thymine (unmethylated) at its position. The major limitation of the bisulfite treatment is its conversion efficiency, as a not complete one causes bias in results. Although, many commercialized kits for bisulfite conversion are available and ensure a minimum of 99% conversion efficacy in most cases (Hernandez *et al.*, 2013; Leontiou *et al.*, 2015). Further details on the bisulfite conversion are provided in section 4.4.1.1 “Bisulfite conversion” at page 125.

REGION-SPECIFIC ANALYSIS

Techniques to estimate methylation of a specific locus generally use PCR to amplify the region of interest.

The **bisulfite sequencing PCR (BSP)** is considered as the gold standard assay to map and quantify region-specific methylation since the 2000s. It consists of converting the DNA with bisulfite, amplifying a region of interest by PCR, and sequencing either the PCR products directly or sequencing several individually cloned PCR products, by Sanger sequencing, to quantify the C/T polymorphisms (Frommer *et al.*, 1992). A complete and detailed description of its principle, process, design, limits, and bias can be found in section 4.4 “Bisulfite sequencing PCR (BSP)” at page 123.

Developed by Herman *et al.* (1996), the **methylation-specific PCR (MSP)** assay is based on specific primers for unmethylated and methylated DNA. The bisulfite converted DNA region is amplified in two PCR reactions using two sets of primers containing both CpG in their sequence but each one with different methylation status. One primer pair contains unmethylated CpG (TG dinucleotides) while the other one contains methylated CpG (CG dinucleotides) in the sequence. The qualitative comparison of the two amplification reactions, by visualization of PCR products on a gel, reveals the relative proportion of methylated DNA compared to unmethylated DNA. Derived from this method, the **MethylQuant** technique and the **sensitive melting analysis after real-time - methylation-specific PCR (SMART-MSP)** technique provide quantitative measurements of methylation proportions using the fluorescence-based real-time PCR (Kristensen *et al.*, 2008; Thomassin *et al.*, 2004). MSP is rapid and accessible, but is not very sensitive as small differences between two samples cannot be measured, and only the methylation proportion of a few CpG sites, the ones covered by primers, is measured.

The **combined bisulfite restriction analysis (COBRA)** assay combines both the bisulfite conversion of DNA and restriction enzyme digestion of PCR products to estimate the degree of methylation at specific loci (Xiong and Laird, 1997). After PCR amplification of the bisulfite converted DNA region of interest, PCR products are digested with specific restriction enzymes (*Bst*UI: CG↓CG,

	GLOBAL MEASUREMENT OF DNA METHYLATION	REGION-SPECIFIC ANALYSIS	ARRAY-BASED ANALYSIS	WIDE PROFILING ANALYSIS
BISULFITE CONVERSION	<ul style="list-style-type: none"> • Alu/LINE-1 assay • Chloroacetaldehyde assay 	<ul style="list-style-type: none"> • BSP • MSP • MS-SNuPE • COBRA • MS-DGGE • MS-SSCA • MethylLight • MS-MCA • MS-DHPLC • Pyrosequencing (PyroMeth) • MethylQuant • EpiTYPER • MassARRAY • MS-HRM • MS-FLAG • SMART-MSP 	<ul style="list-style-type: none"> • MSO • BiMP • GoldenGate • Infinium 	<ul style="list-style-type: none"> • RRBS • WGBS • BC-seq • BSPP • oxBS-seq • MREBS
ENZYME DIGESTION	<ul style="list-style-type: none"> • HPLC • TLC • HPCE • LC-MS • LUMA 	<ul style="list-style-type: none"> • MSRE-PCR • MS-AP-PCR • MS-RDA • MCA-RDA • AIMS • MS-MLPA • MethylScreen 	<ul style="list-style-type: none"> • DMH • RLGS • MCAM • HELP • MethylScope • Mmass • CHARM 	<ul style="list-style-type: none"> • Methyl-seq • DREAM • HELP-seq • MSCC • MRE-seq
AFFINITY ENRICHMENT	<ul style="list-style-type: none"> • 5mC ELISA 	<ul style="list-style-type: none"> • MeDIP-PCR • MIRA 	<ul style="list-style-type: none"> • MeDIP-chip • MIRA-chip 	<ul style="list-style-type: none"> • MeDIP-seq • MIRA-seq • MethylCap-seq

TABLE 4 LIST OF TECHNIQUES TO EVALUATE DNA METHYLATION. (Dahl and Guldborg, 2003; Dhingra *et al.*, 2014; Laird, 2010; Mansego *et al.*, 2013) **5mC ELISA**= 5-methylcytosine Enzyme-Linked Immunosorbent Assay (Deobagkar *et al.*, 1986); **AIMS**= Amplification of Inter-Methylated Sites (Frigola *et al.*, 2002); **Alu/LINE-1 assay**: Alu and LINE-1 repetitive DNA elements assay (Yang *et al.*, 2004); **BC-seq**= Bisulfite conversion Capture - sequencing (Hodges *et al.*, 2009); **BiMP**= Bisulfite Methylation Profiling (Reinders *et al.*, 2008); **BSP**= Bisulfite Sequencing PCR (Frommer *et al.*, 1992); **BSPP**= Bisulfite Padlock Probes (Deng *et al.*, 2009); **CHARM**= Comprehensive High-throughput Arrays for Relative Methylation (Irizarry *et al.*, 2008); **Chloroacetaldehyde assay**: 5mC fluorescent assay (Oakeley, 1999); **COBRA**= Combined Bisulfite Restriction Analysis (Xiong and Laird, 1997); **DMH**= Differential Methylation Hybridization (Huang *et al.*, 1999); **DREAM**= Digital Restriction Enzyme Analysis of Methylation (Jelinek *et al.*, 2009); **EpiTYPER MassARRAY**: MALDI-TOF mass spectrometry-based bisulfite sequencing (Ehrich *et al.*, 2005); **GoldenGate**: Illumina GoldenGate technology applied to methylation profiling (Bibikova and Fan, 2009); **HELP**= Hpal tiny fragment Enrichment by Ligation-mediated PCR (Khuman *et al.*, 2006); **HELP-seq**= Hpal-tiny fragment Enrichment by Ligation-mediated PCR - sequencing (Oda *et al.*, 2009); **HPCE**= High-Performance Capillary Electrophoresis (Fraga *et al.*, 2002); **HPLC**= High-Performance Liquid Chromatography (Kuo *et al.*, 1980); **Infinium**: Illumina Infinium technology applied to methylation profiling (Bibikova *et al.*, 2009); **LC-MS**= Liquid Chromatography - Mass Spectrometry (Friso *et al.*, 2002); **LUMA**= Luminometric Methylation Assay (Karimi *et al.*, 2006); **MCA-RDA**= Methylated CpG island Amplification with Representational Difference Analysis (Toyota *et al.*, 1999); **MCAM**= Methylated CpG island Amplification Microarray (Estécio *et al.*, 2007); **MeDIP-chip**= Methylated DNA Immunoprecipitation - chip (Zhang *et al.*, 2006); **MeDIP-PCR**= Methylated DNA Immunoprecipitation (or mDIP= methylated DNA Immunoprecipitation, or mCIP= methylcytosine Immunoprecipitation) - PCR (Weber *et al.*, 2005); **MeDIP-seq**= Methylated DNA Immunoprecipitation - sequencing (Down *et al.*, 2008); **Methyl-seq**: Methyl-sensitive restriction enzyme - sequencing (Brunner *et al.*, 2009); **MethylCap-seq**: Methyl-DNA binding domain (MBD) capture - sequencing (Brinkman *et al.*, 2010); **MethylLight**: Methylation-specific fluorescent-based real-time PCR (Eads *et al.*, 2000); **MethylQuant**: Discriminative primers based real-time PCR (Thomassin *et al.*, 2004) **MethylScope**: methylation-dependent restriction enzyme microarray (Ordway *et al.*, 2006); **MethylScreen**: methylation-sensitive and methylation-dependent restriction enzyme PCR (Holemon *et al.*, 2007); **MIRA**= Methylated CpG Island Recovery Assay (Rauch and Pfeifer, 2005); **MIRA-chip**= Methylated CpG Island Recovery Assay - chip (Rauch *et al.*, 2007); **MIRA-seq**= Methylated CpG Island Recovery Assay - sequencing (Choi *et al.*, 2010); **MMASS**= Microarray-based Methylation Assessment of Single Samples (Ibrahim *et al.*, 2006); **MRE-seq**= Methylation-sensitive Restriction Enzyme - sequencing (Maunakea *et al.*, 2010); **MREBS**= Methylation-sensitive Restriction Enzyme Sequencing (Bonora *et al.*, 2019); **MS-AP-PCR**= Methylation-Sensitive Arbitrarily Primed PCR (Gonzalogo *et al.*, 1997); **MS-DGGE**= Methylation-Specific - Denaturing Gradient Gel Electrophoresis (Aggerholm *et al.*, 1999); **MS-DHPLC**= Methylation-Specific - Denaturing High Performance Liquid Chromatography (Baumer *et al.*, 2001); **MS-FLAG**= Methylation-Specific - Fluorescent Amplicon Generation (Bonanno *et al.*, 2007); **MS-HRM**= Methylation-Sensitive - High Resolution Melting (Wojdacz and Dobrovic, 2007); **MS-MCA**= Methylation-Specific - Melting Curve Analysis (Worm *et al.*, 2001); **MS-MLPA**= Methylation-Specific - Multiplex Ligation-dependent Probe Amplification (Nygren *et al.*, 2005); **MS-RDA**= Methylation-Sensitive - Representational Difference Analysis (Ushijima *et al.*, 1997); **MS-SNuPE**= Methylation-Sensitive - Single Nucleotide Primer Extension (Gonzalogo and Jones, 1997); **MS-SSCA**= Methylation-Sensitive - Single-Strand Conformation Analysis (Maekawa *et al.*, 1999); **MSCC**= Methylation Sensitive Cut Counting (Ball *et al.*, 2009); **MSO**= Methylation-Specific Oligonucleotide microarray (Gitan *et al.*, 2002); **MSP**= Methylation-Specific PCR (Herman *et al.*, 1996); **MSRE-PCR**= Methylation-Sensitive Restriction Enzyme - PCR (Singer-Sam *et al.*, 1990); **oxBS-seq**= oxidative Bisulfite - sequencing (Booth *et al.*, 2012); **Pyrosequencing (PyroMeth)**: bisulfite conversion and pyrosequencing (Uhlmann *et al.*, 2002); **RLGS**= Restriction Landmark Genomic Scanning (Costello *et al.*, 2000); **RRBS**= Reduced Representation Bisulfite Sequencing (Meissner *et al.*, 2005); **SMART-MSP**= Sensitive Melting Analysis after Real Time - Methylation-Specific PCR (Kristensen *et al.*, 2008); **TLC**= Thin-Layer Chromatography (Schmitt *et al.*, 1997); **WGBS**= Whole Genome Bisulfite Sequencing (or BS-seq, or WGSBS) (Cokus *et al.*, 2008).

TaqI: T↓CGA), cutting specifically CG dinucleotides which are present only at originally methylated sites. The gel electrophoresis of digested products allows the determination of the target sequence methylation level. The main limit of this technique is the use of restriction enzymes with specific sequences narrowing the investigation to specific CpG sites.

The **MethyLight** assay relies on the Taqman technology for real-time PCR amplification by combining methylation-specific priming and methylation-specific fluorescent probing (Eads *et al.*, 2000). Several experimental designs can be found for MethyLight experiments, but the most common one consists of an methylation-independent amplification using two primers and a dual-labeled fluorogenic probe specific to methylated DNA, containing a 5' fluorescent reporter dye and a 3' quencher dye. When the probe is annealed to methylated DNA, the exonuclease activity of the *Taq* DNA polymerase cleaves the probe which releases its 5' reporter fluorescence. Therefore, the fluorescence signal is proportional to the quantity of PCR products from the originally methylated DNA. This method is quantitative and highly sensitive but is expensive, compared to BSP for example, as it requires expensive hybridization probes, and is not reliable to detect heterogeneous methylation on the target sequence (Chatterjee *et al.*, 2017; Dahl and Guldborg, 2003).

The **pyrosequencing** of PCR products from bisulfite converted DNA can be also used to find differential methylation of a specific region. This technique was first used by Uhlmann *et al.* (2002) and is also called **PyroMeth**. Pyrosequencing is a real-time sequencing-by-synthesis technology, based on the luminescence detection from the released pyrophosphate (PPi) on nucleotide incorporation into the complementary strand. Therefore, it is used to evaluate the proportions of C/T polymorphisms at CpG sites (Colella *et al.*, 2003). The pyrosequencing has a high resolution, is quantitative, and does not require cloning as PCR products are directly analyzed, thereby it is a major alternative to BSP. However, it is limited to short sequences, around 150 bp, so the methylation of only a few CpG sites can be quantified. Additionally, pyrosequencing is not expensive but does require access to a pyrosequencer, which can be a limitation for laboratories (Reed *et al.*, 2010).

Among those described bisulfite-based region-specific techniques, only BSP and pyrosequencing can provide a full methylation profile over several CpG sites with a single-nucleotide resolution, while the others quantify the global methylation level of the sequence composed of several CpG sites.

ARRAY-BASED ANALYSIS

Array-based assays use a fixed number of probes, placed on a multiplexed chip, for high-throughput screening of specific loci across the genome.

GoldenGate and **Infinium** are microarray-based sequencing techniques developed by Illumina, originally developed for single nucleotide polymorphisms (SNPs) genotyping at specific loci and they were adapted to specifically detect the C/T polymorphisms at selected CpG sites for methy-

lation profiling.

Several bisulfite-based microarray-based DNA methylation profiling assays have been commercialized by Illumina such as the “GoldenGate Assay For Methylation” (1.5k array targets), “HumanMethylation27 BeadChip” (27k array targets), “Infinium HumanMethylation450 BeadChip” (450k array targets), and the “Infinium MethylationEPIC BeadChip” (850k array targets).

The **GoldenGate** technology applied to methylation profiling is a microbead-based array platform and is based on the specific extension and ligation of correctly hybridized probes on the bisulfite-converted DNA (Bibikova and Fan, 2009). For each CpG site, two sets of two probes are used containing: ① a locus-specific oligo (LSO) that hybridizes the target regions and ② an allele-specific oligo (ASO) that specifically binds with either the methylated allele or the unmethylated one. The allele-specific extension occurs only from the ASO to the LSO and their ligation results in a PCR template. Moreover, the LSO is composed of three parts: ① a target-specific sequence, ② a unique address sequence, and ③ another universal primer sequence, and the ASO is composed of two parts: ① a target/allele-specific sequence and ② a universal primer sequence (two different primer sequences, one for each allele, matching two fluorescently labeled primers). With these universal primers, allele-specific amplification can be carried out. And, the address sequence allows the labeled products hybridization to its complementary probes, coating universal microarray beads, for allele- and target-specific fluorescence reading (Bibikova and Fan, 2009). The standard panel “GoldenGate Methylation Cancer Panel I” spans 1,505 CpG sites from 807 genes (GoldenGate® Assay for Methylation and BeadArray™ Technology, 2010).

The **Infinium** is another microarray technology applied to methylation profiling (Bibikova *et al.*, 2009). Thousands of silica microbeads are placed on the surface of the array and each bead is coated with multiple copies of a 50 bp probe targeting a specific locus. The genomic DNA fragments hybridize with their complementary probes. Then, during the single-base extension step, one of four labeled terminating nucleotides — dideoxynucleotide triphosphates (ddNTPs) — is incorporated. The A and T bases are labeled with 2,4-dinitrophenol (DNP) and the C and G bases with biotin. Next, anti-DNP antibodies and streptavidin labeled molecules specifically bind to the labeled probes to amplify the signal for imaging and fluorescence reading. Two types of the Infinium assay exist, I and II. The Infinium I assay requires two beads per CpG site, one with a probe specific to the methylated sequence and one with a probe specific to the unmethylated sequence. The single-base extension and incorporation of a labeled nucleotide is allele-specific, thereby the presence of a fluorescence signal reveals the methylation state. The Infinium II assay utilizes only one bead with an allele-independent probe. The methylation state is directly determined by the type of fluorescence of the incorporated base as C/T and A/G bases have different labels (Illumina Methylation BeadChips Achieve Breadth of Coverage Using 2 Infinium Chemistries, 2015).

WIDE PROFILING (NGS-BASED) ANALYSIS

The NGS, or second-generation sequencing, consists of the simultaneous sequencing of billions of DNA fragments, giving billions of individual reads that are gathered in contigs — contiguous pieces of the genome — and aligned to the reference genome by bioinformatic analysis. As the bases are sequenced multiples times in several individual reads, it provides a high-depth analysis and thereby highly accurate data (Behjati and Tarpey, 2013; Slatko *et al.*, 2018). The most known NGS technologies are Illumina and Ion Torrent. Its use allows a much higher scale genome analysis compared to Sanger sequencing or arrays with a limited number of targets, but as a downside, it is quite expensive and requires a much more complex bioinformatic analysis to interpret the large amount of data generated.

The **whole-genome bisulfite sequencing (WGBS)** and **reduced representation bisulfite sequencing (RRBS)** are the most known bisulfite sequencing (BS-seq) assays, based on NGS after bisulfite conversion of genomic DNA, to analyze genome-wide methylation profiles on a single nucleotide level. The WGBS is basically a whole-genome sequencing on bisulfite converted DNA and therefore consists of the sequencing of the entire genome, while the RRBS aims to sequence only CG-enriched regions (Cokus *et al.*, 2008; Meissner *et al.*, 2005). To do so, an additional step of DNA fragments size separation (40-220 bp) is performed after digestion by the MspI enzyme (C↓CGG, methylation independent) (Meissner *et al.*, 2008). This technique was developed to overcome the WGBS main limitations, as fewer reads are sequenced it is less expensive and generates fewer data, but still keeps the most interesting regions to analyze (Kurdyukov and Bullock, 2016). The RRBS generally covers around 10% of all CpGs (~2-3 millions of CpG sites) and around 85% of all CGIs (~23,000 CGIs), in the human genome (Smith *et al.*, 2009).

A more recent method, called **methylation-sensitive restriction enzyme bisulfite sequencing (MREBS)** follows the same principle as RRBS, as it combines methylation sensitive restriction enzyme (MSRE) digestion and bisulfite sequencing (addition of a bisulfite conversion step to MRE-seq, methylation-sensitive restriction enzyme sequencing), but expands the CpG coverage (Bonora *et al.*, 2019).

The MREBS protocol uses three methylation-specific restriction enzymes, HpaII (C↓CGG), HinfI (G↓CGC), and AclI (C↓CGC) to cleave unmethylated CpG sites in parallel and a size separation of fragments (50-300 bp) step as RRBS. Moreover, the addition of a bisulfite conversion step allows the determination of methylation levels of the other CpG sites outside the enzyme restriction sites. Thereby, this assay combines the reduced representation towards hypomethylated regions by using MSRE digestion and measurement of adjacent CpG methylation levels by using bisulfite conversion. Several models of methylation analysis using MREBS data have been tested by Bonora *et al.* (2019) to improve both coverage and accuracy. Compared to RRBS and WGBS, MREBS provides a higher coverage of CpG sites than RRBS (~60% of CpG sites) approaching the WGBS coverage

(~75% of CpG sites) while keeping the benefit of a reduced cost compared to WGBS.

4.2.2 ENZYME DIGESTION-BASED ASSAYS

Several assays of DNA methylation quantification are based on the ability of restriction enzymes to cut DNA in a methylation-dependent manner. These MSRE are isoschizomers of restriction endonucleases with a different sensitivity to methylcytosines, as an example both HpaII and MspI cleave C↓CGG sites, MspI can cleave it indifferently of the methylated state of the second cytosine, while HpaII can not cleave it when methylated. Major drawbacks of the use of MSRE are the cleavage site limitation and incomplete digestion leading to a bias towards the methylated state (Laird, 2010).

REGION-SPECIFIC ANALYSIS

The most straightforward technique to evaluate CpG methylation using MSRE is the **MSRE-PCR** (Singer-Sam *et al.*, 1990). It consists of the methylation-specific enzymatic digestion of DNA and PCR amplification of the remaining DNA fragments. Thereby, the absence of PCR products reveals the unmethylation of the target sequence while the presence of them reveals its methylation. In the case of several restriction sites in the target sequence, only one unmethylated site recognized and cleaved by the enzyme is enough to consider the sequence unmethylated. The MSRE-PCR gives an estimation of methylation over several CpG sites and therefore is comparable to the MSP, COBRA, or MethyLight bisulfite-based assays in terms of resolution (Melnikov *et al.*, 2005).

Several other assays using MSRE coupled with other technologies have also been developed. To describe another example, the **MethylScreen** technique uses both MSRE recognizing only unmethylated sites and methylation-dependent restriction enzyme (MDRE) recognizing only methylated sites (McrBC, site: two half-sites RmC within a distance of 40-3,000 bp, cutting between the two half-sites in the proximity of one), coupled with fluorescence-based real-time PCR to quantify methylation (Holemon *et al.*, 2007).

ARRAY-BASED ANALYSIS

The MSRE-associated discrimination of methylated and unmethylated sites can also be combined with microarray analysis. The **HpaII tiny fragment enrichment by ligation-mediated PCR (HELP)** assay is based on the comparison of HpaII (C↓CGG, unmethylated) and MspI (C↓CGG, methylation insensitive) digested DNA fragments. Using random priming, fragments derived from both HpaII or MspI digestions are labeled with a different fluorophore each and then cohybridized to a microarray. The ligation-mediated PCR and fluorescence detection allow to get the HpaII or MspI representations and the calculation of HpaII/MspI log ratios gives an estimation of hypomethylated and hypermethylated loci (Khulan *et al.*, 2006).

The **MethylScope** assay is an extension of the MethylScreen assay, also using MDRE (McrBC) digestion associated with microarray analysis (Ordway *et al.*, 2006).

WIDE PROFILING (NGS-BASED) ANALYSIS

Several assays have been adapted to incorporate an NGS analysis for genome-scale methylation analysis. The MSRE-PCR gave the **MRE-seq**, consisting in MSRE digestion followed by sequencing (Maunakea *et al.*, 2010). And, the HELP has been extended to **HELP-seq**, following the same principle but using NGS instead of a microarray analysis (Oda *et al.*, 2009).

4.2.3 AFFINITY ENRICHMENT-BASED ASSAYS

The affinity enrichment of methylated regions is based on 5mC antibodies or methyl-binding proteins (Laird, 2010). Two main techniques have been developed, the **methylated DNA immunoprecipitation (MeDIP)** (or mDIP= methylated DNA immunoprecipitation, or mCIP= methylcytosine immunoprecipitation) and the **methylated CpG island recovery assay (MIRA)**. The MeDIP is a DNA purification technique by immunoprecipitation using an antibody specific to 5mC (Weber *et al.*, 2005). The MIRA technique relies on the enrichment of methylated DNA regions using methyl-binding proteins, such as methyl-CpG-binding protein 2 (MeCP2), methyl-CpG-binding domain protein 2 (MBD2), or methyl-CpG-binding domain protein 3 like 1 (MBD3L1) (Choi *et al.*, 2010; Rauch and Pfeifer, 2005).

Both MeDIP and MIRA principles can be used followed by ① PCR amplification: **MeDIP / MeDIP-PCR** and **MIRA**, ② microarray: **MeDIP-chip** and **MIRA-chip**, or ③ NGS: **MeDIP-seq** and **MIRA-seq**, depending on the desired scale of purified methylated DNA fragments detection (Choi *et al.*, 2010; Down *et al.*, 2008; Rauch and Pfeifer, 2005; Rauch *et al.*, 2007; Weber *et al.*, 2005; Zhang *et al.*, 2006). The affinity enrichment-based assays do not rely on either bisulfite conversion or restriction sites, but can not provide single-CpG resolution methylation profiles.

KEY POINTS

- ➔ A plethora of methods to evaluate the CpG methylation have been developed over the years, a non-exhaustive classified list is presented in [Table 4](#).
- ➔ The DNA methylation assays can be classified into three major groups depending on the principle used to discriminate 5mC: ① bisulfite-based assays, ② enzyme digestion-based assays and ③ affinity enrichment-based assays. Another classification based on their coverage can also be proposed: ① global measurement of DNA methylation, ② the region-specific assessment of methylation, ③ array-based analysis and ④ wide profiling or genome-scale analysis.
- ➔ The bisulfite-based techniques discriminate methylated cytosines based on their ability to be unaffected by the sodium bisulfite treatment which converts unmethylated ones into uracils.
- ➔ The enzyme-based techniques discriminate methylated cytosines based on the sensitivity of restriction enzyme towards methylated restriction sites, unrecognized sites that can not be cleaved when methylated.
- ➔ The affinity-based techniques discriminate methylated cytosines using either 5mC antibodies for methylated DNA immunoprecipitation or methyl-binding proteins for methylated regions enrichment.

^ [Back to Table of Contents](#)

4.3 BISULFITE SEQUENCING PCR (BSP)

Historically, the deamination of cytosines by sodium bisulfite was described in the 1970s ([Hayatsu *et al.*, 1970](#); [Shapiro *et al.*, 1973](#); [Shapiro *et al.*, 1970](#); [Wang *et al.*, 1980](#)) and with the emergence and commercialization of region-specific genomic sequencing technique in the 1980s ([Church and Gilbert, 1984](#); [Sanger *et al.*, 1977](#); [Smith *et al.*, 1985](#); [Smith *et al.*, 1986](#)), the combination of both gives rise to a novel method to study DNA methylation in a specific genomic region, called Bisulfite Genomic Sequencing. This method was used for the first time in the 1990s by [Frommer *et al.* \(1992\)](#) to identify and map 5-methylcytosines (5mC) in genomic DNA. The process consists of a DNA bisulfite conversion, polymerase chain reaction (PCR) amplification of a target region, and sequencing of either PCR products to get an average of the molecules population or individual clones to map methylation status within a single DNA molecule ([Clark *et al.*, 1994](#); [Frommer *et al.*, 1992](#); [Rein *et al.*, 1998](#); [Yuanxiang *et al.*, 1997](#)). Then, the technique was improved to quantify the methylation of cytosines in a population of DNA molecules ([Lewin *et al.*, 2004](#); [Paul and Clark, 1996](#); [Suzuki *et al.*, 2000](#)).

This method was called for the first time bisulfite sequencing PCR (BSP) in opposition to methylation-specific PCR (MSP) in [Li and Dahiya \(2002\)](#), the publication introducing the MethPrimer program to design primers for bisulfite-based PCR methods. Although, in some publications, it can be still referred to as bisulfite genomic sequencing, or even bisulfite sequencing, but it must not be mistaken for bisulfite sequencing techniques using NGS sequencing ([Li and Tollefsbol, 2011](#); [Lizardi](#)

et al., 2016). The bisulfite sequencing PCR (BSP) name describes the three main steps of the technique : ① the **bisulfite** conversion, ② the **PCR** amplification, and ③ the **sequencing** of either amplicons or clones (Figure 24). This way, it can be distinguished from other techniques such as MSP, bisulfite pyrosequencing, or WGBS.

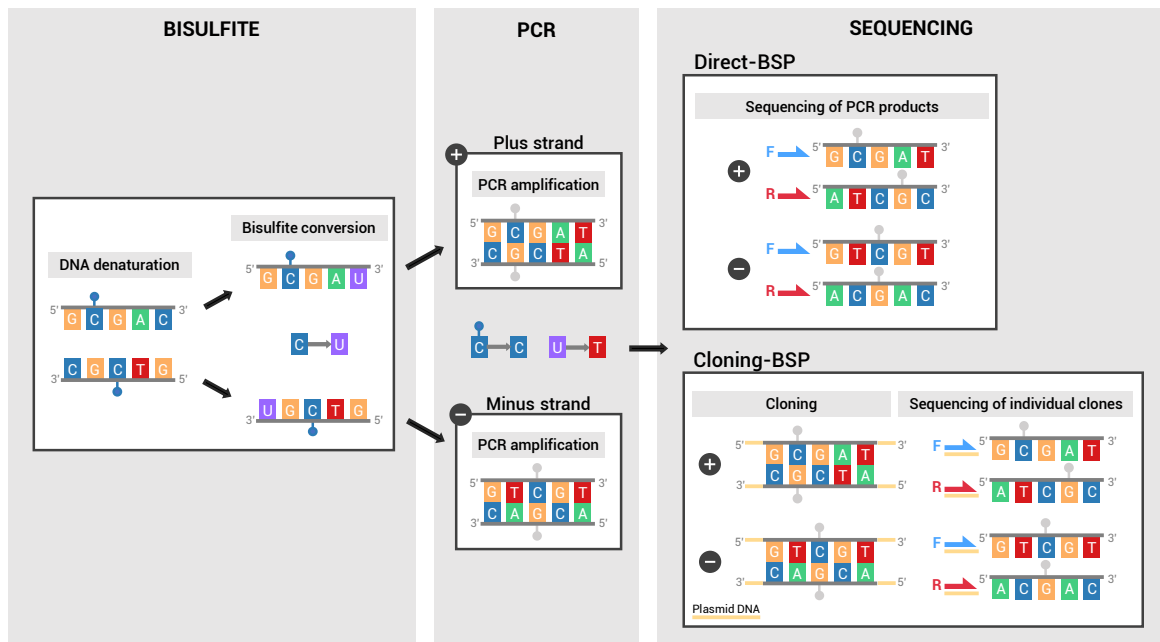


FIGURE 24 BISULFITE SEQUENCING PCR (BSP) EXPERIMENTAL PRINCIPLE.

The BSP technique was a gold standard method to estimate methylation of a specific locus in the 2000s-2010s, mainly due to the standardization and commercialization of bisulfite kits and the growing accessibility of Sanger sequencing (Hernandez *et al.*, 2013; Leontiou *et al.*, 2015; Li and Tollefsbol, 2011; Lizardi *et al.*, 2016). Despite the development of bisulfite pyrosequencing, which provides a quantitative measurement of DNA methylation without the cloning requirement, the BSP technique remains more accessible since bisulfite pyrosequencing requires a pyrosequencer (Reed *et al.*, 2010).

Nowadays, a plethora of methods are used to study DNA methylation depending on experiment requisites (Kurdyukov and Bullock, 2016; Pajares *et al.*, 2021), but BSP is still widely used for region-specific studies (Akika *et al.*, 2017), for example as a first approach for large studies because of its cost-effectiveness compared to NGS based techniques (Ismail *et al.*, 2020; Liu *et al.*, 2021b; Moschny *et al.*, 2020) or for locus-specific confirmation of WGBS or RRBS results (Fan *et al.*, 2020; Zhang *et al.*, 2017b; Zhu *et al.*, 2019).

4.4.1 EXPERIMENTAL PROCESS

The BSP experimental process is composed of three main steps : bisulfite conversion, polymerase chain reaction (PCR), and sequencing (Figure 24). For its last step, two alternative ap-

proaches for sequencing exist depending on the addition or not of a cloning step before sequencing, called respectively cloning-BSP or direct-BSP. This variation within the process will result in different analysis strategies (see section 4.4.3 “Analysis strategy and tools” at page 130).

BISULFITE CONVERSION

The first key step of BSP is the bisulfite conversion of the extracted DNA molecules by a sodium bisulfite treatment. The sodium bisulfite NaHSO_3 comes from the dilution of sodium metabisulfite $\text{Na}_2\text{S}_2\text{O}_5$ in water. It is composed of bisulfite HSO_3^- and sodium Na^+ and can mediate the deamination reaction of cytosine nucleotides to form uracil nucleotides. This reaction was first described by Hayatsu *et al.* (1970). As depicted in Figure 25, the bisulfite conversion of cytosine into uracil proceeds in 3 steps (Hatakeyama *et al.*, 2013; Hayatsu *et al.*, 1970):

1. Sulfonation at the C6 position of the cytosine residue: cytosine [C] to cytosine sulfonate [C-SO_3^-].
2. Hydrolytic deamination at the C4 position: cytosine sulfonate [C-SO_3^-] to uracil sulfonate [U-SO_3^-].
3. Alkaline desulfonation: uracil sulfonate [U-SO_3^-] to uracil [U].

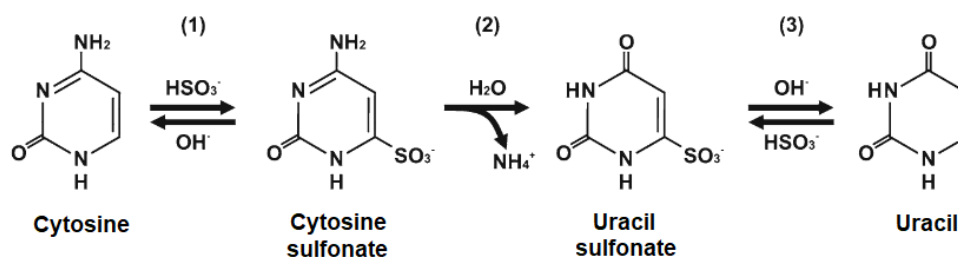


FIGURE 25 BISULFITE-MEDIATED DEAMINATION OF CYTOSINE. (1) Sulfonation of the cytosine to cytosine sulfonate. (2) Hydrolytic deamination of the cytosine sulfonate to uracil sulfonate. (3) Alkaline desulfonation of the uracil sulfonate to uracil. Figure from Hatakeyama *et al.* (2013).

When the deamination of cytosine by bisulfite was described in the 1970s, the 5-methylcytosine was considered a minor constituent of DNA, and the role of cytosine methylation was not yet established (Hayatsu *et al.*, 1970). But, it was already known that the bisulfite deamination of cytosines and 5-methylcytosines have different rates: the 5-methylcytosine deamination by bisulfite treatment is much slower than for cytosine. As the technology and knowledge advance, Frommer *et al.* (1992) were the first to use this difference as a way to determine the localization of 5-methylcytosines, as cytosines are converted to uracils while 5-methylcytosines are not, thus it creates a difference in the DNA sequence.

Before the sodium bisulfite treatment, as cytosine residues have to be exposed to bisulfite, the DNA must be fully denatured, in the single strand state. After the conversion of cytosines, uracils on one strand cannot pair with guanines on the opposite strand, therefore the two converted strands are no longer complementary and must remain single-stranded until use.

The bisulfite reaction has two major limitations. First, the conditions of the reaction have to

be monitored to evaluate the conversion efficiency. As explained further below, in section “Bisulfite conversion efficiency” at page 139, if the conversion is incomplete some cytosines remain unaffected and if the treatment is incubated for too long, the bisulfite can start to deaminate 5-methylcytosines (Hayatsu, 2008a; Hayatsu, 2008b). Secondly, the bisulfite treatment causes strand breakage in DNA due to the formation of the $\text{SO}_3^{\cdot-}$ radical (Hayatsu and Miller, 1972). This radical-mediated degradation of DNA is reduced by the addition of hydroquinone, a radical scavenger, in the bisulfite treatment.

POLYMERASE CHAIN REACTION (PCR)

The second step of BSP is to selectively amplify the targeted DNA region to get enough material for cloning or sequencing. As routinely performed in laboratories, the DNA is amplified by polymerase chain reaction (PCR).

The PCR is based on the hybridization of oligonucleotides, complementary to the targeted region, which serves as primers for the polymerase to regenerate the complementary strand, and therefore it produces a new copy of the DNA fragment. Briefly, a PCR is classically composed of 25 to 50 cycles of 3 steps : ① DNA denaturation, temperature around 95°C, ② Annealing of primers on their specific DNA complementary sequence, temperature around 50-65°C depending on primer melting temperature (T_m) and ③ Elongation for the new DNA strand to be synthesized by the DNA polymerase, temperature for optimum DNA polymerase activity generally at 72°C (Green and Sambrook, 2019b).

In the context of bisulfite-treated DNA amplification, many limitations come from the modifications of DNA by bisulfite (single-strand DNA with no more complementary strand, loss of base heterogeneity due to the T redundancy, DNA degradation...) that need to be taken into consideration for primer design, a crucial step to ensure a specific, efficient and unbiased amplification. As bisulfite-converted DNA strands are no longer complementary, the PCR primers have to be designed using one of those two templates. Therefore, for the same DNA sequence, sequencing results will be different depending on the chosen amplified strand, so it needs to be included in the analytic process (Figure 24). Primer design guidelines for BSP are depicted in Appendix 1 “Primer design for bisulfite sequencing PCR” at page 305 and the potential PCR biased amplification in BSP experiments is explained in section “PCR bias: unequal amplification of methylated and unmethylated DNA” at page 142.

The two major limitations, which are ① primer specificity issues due to reduced heterogeneity and ② low DNA material due to degradation by bisulfite treatment, can be partly resolved by adjusting the PCR protocol.

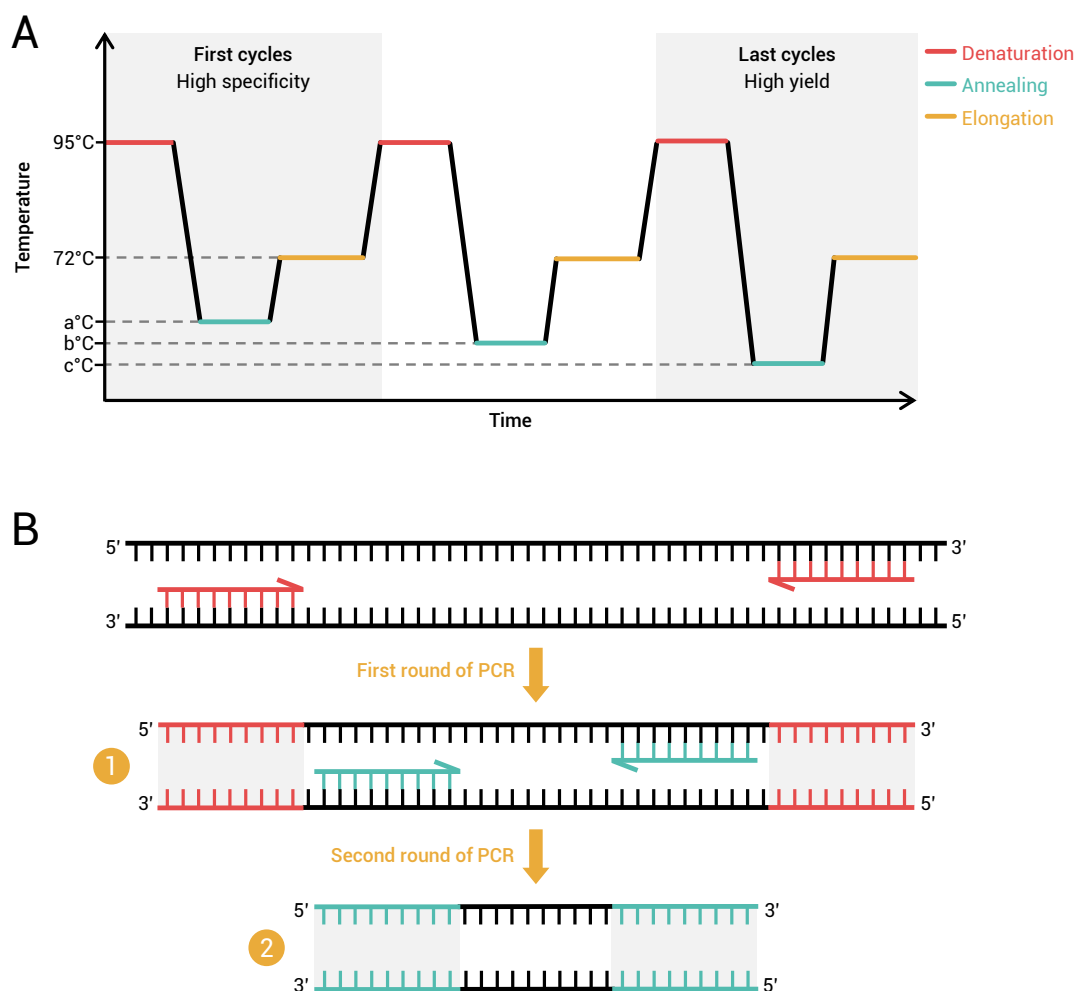


FIGURE 26 SCHEMATIC REPRESENTATION OF TOUCHDOWN AND NESTED PCR PROTOCOLS. These two PCR protocols are used in BSP experiments to improve the yield and specificity of the PCR reaction. **A.** In touchdown PCR protocols, the specific amplicon is highly amplified and becomes predominant so when the annealing temperature (T_a) decreases, the specificity also decreases, but the predominant template out-competes the non-specific ones. The T_a annotated a, b and c are generally comprised between 50°C and 65°C, and spaced by 1 to 5°C. The number of cycles, duration of steps, temperatures for denaturation/elongation/annealing, and range of T_a are varying depending on the chosen protocol, reagents, polymerase, primers T_m , and region of interest. The first steps and last steps of the PCR program outside the cycle steps are not represented. **B.** For nested PCR protocols, two rounds of PCR are performed with two different primer pairs, an outer one in red and an inner one in green. The first amplicon ① is used as a template to amplify the second one ②. By using two rounds of amplification it increases the yield of the PCR product, and by using two sets of primers it increases the specificity towards the targeted region.

Touchdown PCR

The touchdown PCR (TD-PCR) is frequently used as PCR protocol for BSP and relies on a progressive decrease of the annealing temperature (T_a) over the cycles during the PCR reaction (Li and Tollefsbol, 2011; McDonald and Kay, 1997; Nagane *et al.*, 2000; Shen *et al.*, 2007). This protocol ensures both high specificity and a strong yield of the desired PCR product (Figure 26A). The hybridization of primers starts in the first cycles above the optimal annealing temperature to guarantee the strong specificity towards the targeted region as only perfect primer-template hybrids can be formed. Then, as the cycling program advances, the annealing temperature drops progressively to increase the yield of amplification but it does not lead to non-specific products amplification because the predominant specific products out-compete the non-specific ones thanks to the stringent initial cycles (Green and Sambrook, 2018; Hecker and Roux, 1996; Korbie and Mattick, 2008; Roux, 2009).

Nested PCR

Another possible and complementary PCR protocol used for BSP that can improve the yield of the PCR product and reduce non-specific amplification is the nested PCR (Ashapkin *et al.*, 2020; Chen *et al.*, 2017a; Grunau *et al.*, 2001; Li and Tollefsbol, 2011; Lizardi *et al.*, 2016; McDonald and Kay, 1997; Olek *et al.*, 1996). The nested PCR protocol involves two sequential amplification using two different pairs of primers, an outer one and an inner one (Figure 26B). The first amplification produces an amplicon between the outer primers which serves as a template for the second amplification using the inner primers. The double amplification step increases the yield of the amplicon and the use of two distinct sets of primers for the same targeted region improves the specificity of the fragment produced (Green and Sambrook, 2019a; Haff, 1994; Roux, 2009).

CLONING (OPTIONAL)

In the case of cloning-BSP, PCR products are inserted in a plasmid and amplified in transformed bacteria. Bacterial clones are selected, and individually amplified before plasmid extraction. As a competent bacteria assimilate a single plasmid molecule during transformation, all of the plasmids of an individual clone possess a unique sequence of the PCR product inserted. This way, the cloning of PCR products from bisulfite-converted DNA reveals the CpG methylation statuses of a single initial DNA molecule from the sample.

Generally in publications, 5 to 10 clones are sequenced, revealing the methylation status of 5 to 10 initial DNA molecules, which seems to be considered a good compromise between the representativeness of results and the cost and time investment (Li and Tollefsbol, 2011; Lizardi *et al.*, 2016). Yet, this approach is appropriate when differences are really important, but some publications state the necessity to have more clones, around 50 to 100 clones, to obtain a statistically significant estimation of the methylation proportion with good precision and prone the sequencing of PCR products directly for a better representation despite potential additional bias (Mühlisch *et al.*, 2007; Paul

and Clark, 1996; Rohde *et al.*, 2010; Voss *et al.*, 1998).

SEQUENCING

The Sanger sequencing technique was conceived by Sanger *et al.* in 1977 and since then it is widely used to decipher the base sequence of a specific and single DNA molecule. This technology has been improved and automated over the years and nowadays, and despite the emergence of the NGS technology, it remains the most accessible method to determine a DNA sequence (Slatko *et al.*, 2011; Valencia *et al.*, 2013).

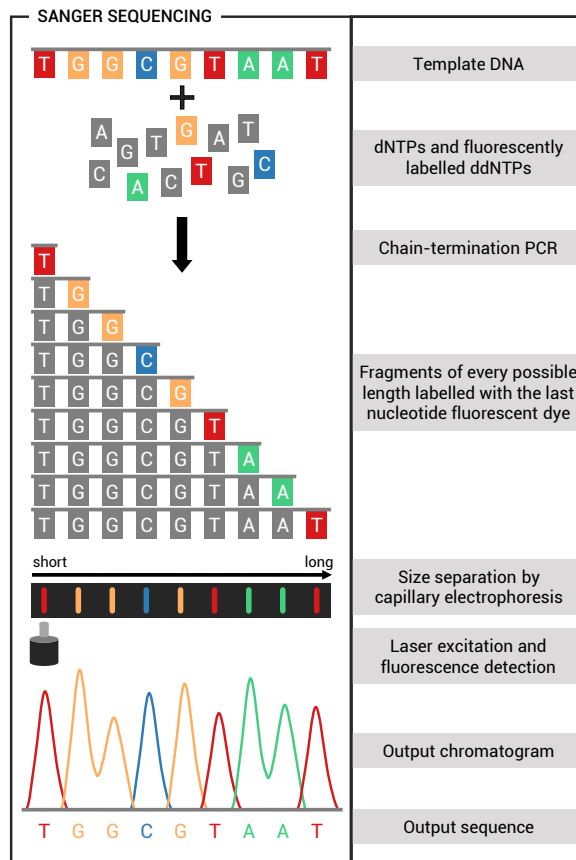


FIGURE 27 AUTOMATED SANGER SEQUENCING PRINCIPLE. The DNA molecule is amplified by chain-termination PCR. The polymerase can incorporate normal nucleotides (deoxynucleotide triphosphates (dNTPs), in grey) or dye-labelled modified nucleotides (ddNTPs, colored). The incorporation of a ddNTP by the DNA polymerase stops DNA synthesis. As the DNA polymerase synthesizes DNA only from 5' to 3', the PCR results in a multitude of fragments, varying in length and labeled with the terminal ddNTP dye. The DNA fragments are size separated by capillary gel electrophoresis. By reading the gel from the smallest fragment to the longest fragment, the fluorescence of the bands reveals the DNA sequence. The result of the gel reading can be seen as a four-dye trace plotted in a chromatogram, where each peak of the fluorescence signal corresponds to a specific nucleotide.

The automated Sanger sequencing relies on a chain-termination PCR to amplify a specific DNA fragment, by incorporating either normal nucleotides (dNTPs) or fluorescently labeled terminating modified nucleotides (ddNTPs) that stops the fragment synthesis (Figure 27). As the polymerization progresses, amplicons at every possible length are synthesized and labeled with a specific fluorescence depending on the last nucleotide. The migration by capillary electrophoresis (CE) separates fragments depending on their sizes, resulting in a string of fluorescence signals.

The automated sequencer reads these four fluorescence signals and generates an electropherogram, visible as a chromatogram, which is a four-dye trace saved as an ABIF (Applied Biosystems, Inc. Format) format file with the .ab1 extension or as an SCF (Staden Chromatogram Files or Sequence Chromatogram File) format file with the .scf extension (Applied Biosystems Genetic Anal-

ysis Data File Format, 2006; Dear and Staden, 1992). And finally, the base caller program assigns bases corresponding to each primary chromatogram peak to get the resulting sequence, saved as a FASTA format file with the .fasta extension (Hyman *et al.*, 2010; Pearson and Lipman, 1988).

Two types of BSP approach exists depending on the template sequenced: ① direct-BSP if the PCR products are directly sequenced and ② cloning-BSP if the PCR products are first cloned in vectors before the sequencing of individual clones (Chatterjee *et al.*, 2017).

KEY POINTS

- The bisulfite sequencing PCR (BSP) is a technique to assess DNA methylation levels of a specific region of interest and is composed of three steps: ① the bisulfite conversion of DNA, ② the PCR amplification of the target region, and (3) the sequencing of either the PCR products directly or PCR products individualized in clones (Figure 24).
- A sodium bisulfite treatment converts unmethylated cytosines into uracils, while methylated ones remain cytosines. Thereby, after PCR amplification the original unmethylated cytosine is revealed by the T base while the unmethylated one by a C base (Figure 25).
- Two types of BSP experiment exist: the direct-BSP one and cloning-BSP one. The first one consists in directly sequencing the mix of PCR products, while the second one consists in cloning the PCR products before individual clones are sequenced.
- Information and guidelines about primer design for BSP experiments are provided in Appendix 1 “Primer design for bisulfite sequencing PCR” at page 305.

↗ [Back to Table of Contents](#)

4.4.2 ANALYSIS STRATEGY AND TOOLS

The two BSP approaches, direct-BSP and cloning-BSP, differ in their analytic process. Indeed, the CpG methylation percentage calculation is not the same either it is based on direct-BSP or cloning-BSP data (Figure 28).

Although PCR products were cloned in the first publications describing the BSP experiment, in Frommer *et al.* (1992) and Clark *et al.* (1994), they mentioned an alternative way by directly sequencing the PCR products. Later on, the quantification of DNA methylation using the direct-BSP method was demonstrated by Lewin *et al.* (2004) and it led to its utilization in numerous studies afterward. These two methods were described in several literature reviews such as: Chatterjee *et al.*, 2017; Hernandez *et al.*, 2013; Martisova *et al.*, 2021; Mikeska *et al.*, 2010; Pajares *et al.*, 2021 as well as in several protocols such as: Ashapkin *et al.* (2020); Li and Tollefsbol (2011); Lizardi *et al.* (2016); Zhang *et al.* (2009).

CLONING-BSP: CLONING OF PCR PRODUCTS AND SEQUENCING OF INDIVIDUAL CLONES PLASMIDS

The cloning-BSP is based on the random separation of unique DNA molecules, with different methylation patterns, in different clones, as a representation of the methylation diversity in the total

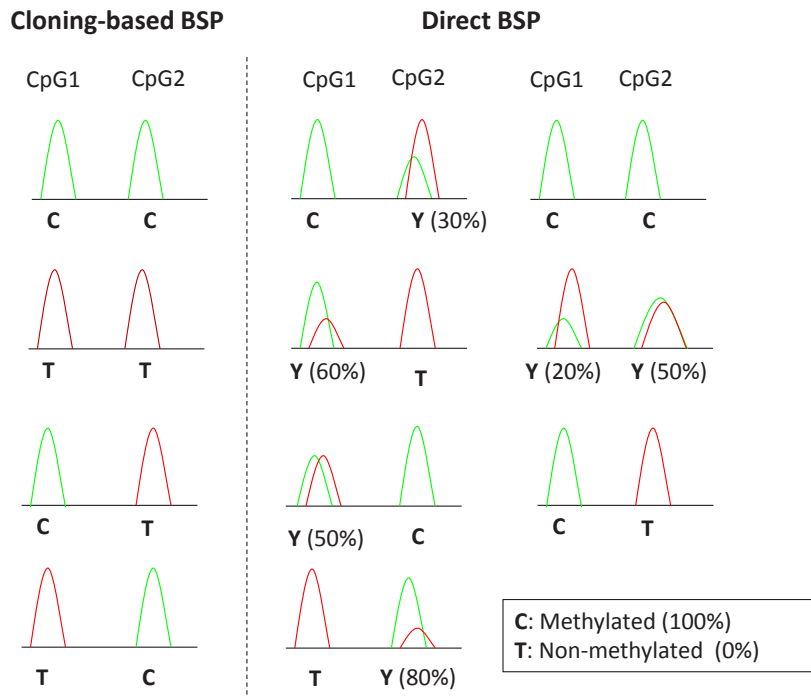


FIGURE 28 COMPARISON OF SEQUENCING TRACE DATA BETWEEN CLONING-BSP AND DIRECT-BSP APPROACHES. The figure illustrates the chromatogram data of possible results from BSP, for two CpG sites. On the left panel, possible results of unique clones from cloning-BSP. On the right panel, possible results from the sequencing of PCR products from direct-BSP. A C indicates a 100% methylated cytosine, a T indicates a 0% methylated cytosine and a Y indicates partial methylation between 0% and 100%. Figure from Hernandez *et al.* (2013).

DNA population. Thereby, the sequencing of several clones provides an estimation of methylated cytosine proportions per CpG site. For each CpG, the number of clones for which a C base (or G for reverse sequencing) is found, over the total of clones sequenced assesses the proportion of methylated cytosines in the original DNA mix (Figure 28). As mentioned previously, in the literature it seems to be admitted that 10 clones are sufficient to reveal 10% differences in methylation, even if several publications recommend having more clones to increase the sensitivity (Mühlisch *et al.*, 2007; Paul and Clark, 1996; Rohde *et al.*, 2010; Voss *et al.*, 1998).

Over the years, several methods have been developed and used to facilitate the analysis of clone methylation status from cloning-BSP. Initially, the results of clone sequencing were manually analyzed by retrieving, on the sequencing gel, the base corresponding to the CpG positions for each clone (Clark *et al.*, 1994; Frommer *et al.*, 1992; Paul and Clark, 1996; Stirzaker *et al.*, 1997). Then, in 2000, Grunau *et al.* developed a software, called *MethTools*, able to align the sequences of clones with the original reference sequence, to deduce cytosine methylation patterns and generate graphical outputs, available on both Linux and Mac operating systems. Later, a new version has been released as a web server available online, named *MethTools 2.0* (available at http://methdb.univ-perp.fr/methtools/MethTools2_submit.html). This new version has the same purpose of analyzing cloning-BSP data, and requires as an input the genomic and subclones sequences in a fasta file. In 2001, Anbazhagan *et al.* published a spreadsheet-based program – a Microsoft Excel file with a built-in program – to identify CpG island (CGI) and to facilitate the calculation of DNA methylation percentages from cloning-BSP data.

More advanced software were then developed, to align, visualize and quantify CpG methy-

lation from subclones sequences, including quality control features, graphical outputs and statistics (Figure 29):

- *BiQ Analyzer* (<https://biq-analyzer.bioinf.mpi-inf.mpg.de/>, Bock *et al.*, 2005)
- *QUMA (Quantification tool for Methylation Analysis)* (<http://quma.cdb.riken.jp/>, Kumaki *et al.*, 2008)
- *BISMA (Bisulfite Sequencing DNA Methylation Analysis)* (<http://services.abc.uni-stuttgart.de/BDPC/BISMA/>, Rohde *et al.*, 2010)

	Feature	QUMA	BiQ Analyzer	BISMA
Requirements and features	The software is freely available	✓	✓	✓
	Installation free usage	✓	-	✓
	Analysis of unique sequences	✓	✓	✓
	Analysis of repetitive sequences	-	-	✓
Data processing and analysis	Uploading of sequencing files in text format	✓	✓	✓
	Uploading of ABI sequencing files	-	-	✓
	Assistance in sequence direction detection	✓	✓	✓
	Highly automated data processing algorithm	✓	-	✓
	Automatic detection of sequence direction	✓	-	✓
	Automatic vector removal	✓	-	✓
	High speed of data processing and analysis	✓	-	✓
	Improved CpG site detection accuracy	-	-	✓
Quality control features	Sequence identity measurement	✓	✓	✓
	Conversion rate measurement	✓	✓	✓
	Detection of insertions/deletions	✓	-	✓
	Filtering N-sites at cytosine positions during analysis	-	-	✓
	Basic detection of clonal molecules	-	✓	✓
	Improved detection of clonal molecules	-	-	✓
Statistics and presentation	Output of an annotated multiple sequence alignment	✓	✓	✓
	CpG site average report requires a minimum number of clones for data processing	-	-	✓
	Sequence sorting according to methylation in all output files	✓	-	✓
	Direct web presentation of final results	-	✓	✓
	Compatibility with BDPC compilation software	-	✓	✓

FIGURE 29 COMPARISON OF FEATURES BETWEEN THE THREE TOOLS ANALYZING CLONING-BSP DATA: QUMA, BIQ ANALYZER, AND BISMA. Figure from Rohde *et al.* (2010), introducing the *BISMA* software.

Additionally, specifically for plants models, the software *CyMATE* and *Kismeth* has been developed to analyze cloning-BSP data, as in plants, cytosines methylation is not limited to CpG, as cytosines can be methylated in several contexts (CG, CHG, and CHH, with H meaning non-G) (Gruntman *et al.*, 2008; Hetzl *et al.*, 2007).

The most recent one, *BISMA*, is the only program to accept .ab1 sequencing files as input, but they are only used to extract the DNA sequence from it. Therefore, all of these tools are specifically designed to analyze the methylation statuses of subclones. They base their analysis process on the base-called sequence obtained from the sequencing run, indicating either a C base or a T base and determining the cytosine methylation statuses at a single-molecule resolution.

DIRECT-BSP: DIRECT SEQUENCING OF PCR PRODUCTS

As all PCR products are directly sequenced in the direct-BSP method, it means that a mix of DNA molecules with different methylation statuses at CpG positions are being sequenced simultaneously (Figure 30) (Jiang *et al.*, 2010; Lewin *et al.*, 2004; Myöhänen *et al.*, 1994; Parrish *et al.*, 2012; Paul and Clark, 1996). Hence, in the sequencing trace, two signals can co-exist at the cytosine position, the one from the base corresponding to the methylated state of the CpG site (C when using a forward primer, G when using a reverse primer) and the one from the base corresponding to the unmethylated state of the CpG site (T when using a forward primer, A when using a reverse primer) (Figure 28 and Figure 30).

The direct-BSP analysis approach is based on the assumption that the signal intensity of the base associated fluorescent dye, detected during sequencing, is proportional to the number of DNA molecules containing this residue, and that signal intensities can be compared between the four bases. It means that the relative quantities of nucleotides within a DNA mix can be estimated using the maximum signal intensity values and that, by comparing them, they reveal the proportion of nucleotides in the DNA pool, for a given position.

Accordingly, the signal intensities for each base, given as electropherogram peak heights, are utilized to compute their proportion in the sequenced DNA molecule mix (Jiang *et al.*, 2010; Parrish *et al.*, 2012). The values for peak heights can be retrieved through free software that can open chromatogram files to view traces, such as *4 Peaks* (only for Mac operating systems, <https://nucleobytes.com/4peaks/>) or *Chromas* (<http://technelysium.com.au/wp/chromas/>). The methylation percentage is calculated based on the following formulas of signals ratios, in which the nucleotide letter refers to the peak height value of the base associated fluorescent signal (Figure 30):

$$\text{Methylation percentage (forward sequencing)} = \frac{\text{peak height of C}}{\text{peak height of C} + \text{peak height of T}} \times 100$$

$$\text{Methylation percentage (reverse sequencing)} = \frac{\text{peak height of G}}{\text{peak height of G} + \text{peak height of A}} \times 100$$

However, this assumption has limitations. Firstly, the incorporation efficiencies of the labeled terminator nucleotides (ddNTPs) can be different from one another. Thus, the proportionality between signal intensities and relative quantities might be biased. Secondly, as each ddNTP is linked to a different fluorochrome, the comparison of different fluorescence signals can also introduce bias in the calculation of proportions between nucleotides. That's why some publications qualify the direct-BSP approach as non-quantitative or semi-quantitative, and that the cloning-BSP approach is often preferred (Chatterjee *et al.*, 2017; Chhibber and Schroeder, 2008; Mikeska *et al.*, 2010; Parker *et al.*, 1995).

Compared to cloning-BSP, the direct-BSP approach presents the advantage of reducing the experiment duration as it skips the cloning step and avoids the multiplication of subclones sequencing

costs (Chatterjee *et al.*, 2017; Martisova *et al.*, 2021; Pajares *et al.*, 2021). It is therefore particularly useful for the assessment of DNA methylation in studies with numerous samples, such as cohorts of patients for example (Liu *et al.*, 2021b; Moschny *et al.*, 2020; Schiele *et al.*, 2021). Additionally, it is relevant to sequence PCR products to get preliminary results before committing to the cloning steps. Thereby, even in cloning-BSP studies, the direct-BSP approach can be performed to obtain preliminary or complementary results (Martisova *et al.*, 2021).

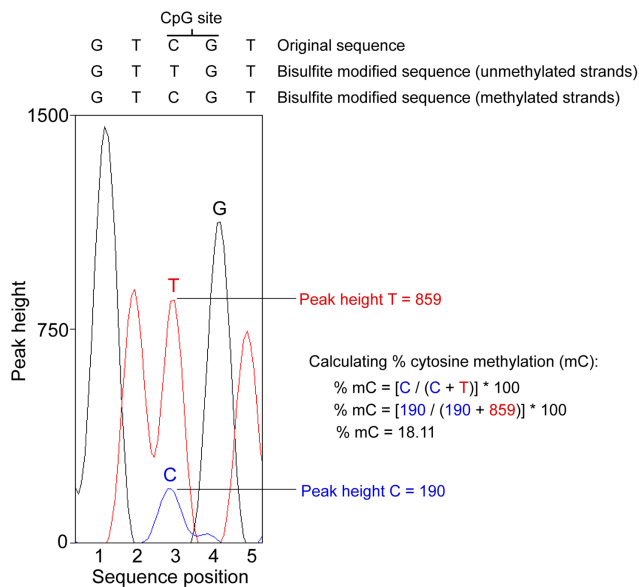


FIGURE 30 METHOD OF METHYLATION LEVEL CALCULATION BASED ON SEQUENCING CHROMATOGRAM FILE FOR DIRECT-BSP. The figure represents an example of methylation level calculation at a single CpG site from the forward sequencing data. At the CpG position, there are two signals detected, one for both the thymine and cytosine bases, corresponding respectively to the unmethylated and methylated state of the cytosine. So these two chromatogram peaks are compared to evaluate the proportion of methylation. The values of peak heights are used to compute the ratio of signals for the methylated (C) state over the total of the methylated (C) and unmethylated (T) states. In the example, as the peak height of the methylated state signal represents 18.11% of the total signals, the methylation level of the CpG is estimated at 18.11%. For the reverse sequencing data, as the sequence corresponds to the reverse complement, methylation levels are determined using the signals of adenine for the unmethylated state and guanine for the unmethylated state. Figure from Parrish *et al.* (2012).

Historically, the direct sequencing of bisulfite converted PCR products to assess CpG methylation was first done by Myöhänen *et al.* (1994) and aimed to improve the BSP technique described by Frommer *et al.* (1992) beforehand. The direct sequencing of amplicons was performed by an automated sequencer with fluorescently labeled primers to determine the CpG methylation status, or for partial methylation to approximately estimate the degree of methylation, with an accuracy of 25%, by visually comparing the chromatogram peaks.

The first use of electropherogram peaks comparison in a quantitative and reproducible manner was realized by Kwok *et al.* (1994). This method was used to quantify the single nucleotide polymorphisms (SNPs) in a DNA pool, to estimate the allele frequencies, based on the peak heights of the chromatogram traces. Although it was done manually in this publication, this analytic method for sin-

gle nucleotide polymorphisms (SNPs) quantification was then automated by Qiu *et al.* (2003) using a similar methodology. And by comparing their method with the pyrosequencing method, they showed that it was more accurate, sensitive and reproducible than pyrosequencing for SNPs quantification.

As the presence of two peaks at CpG positions can be interpreted as SNPs, this method can be applied to cytosine methylation calculation for direct-BSP. Hence, this methodology was first applied to PCR products coming from bisulfite-converted DNA, by Paul and Clark (1996) to quantify cytosine methylation by an automated genomic sequencing approach, specifically developed for this purpose. Paul and Clark considered that cloning-BSP requires 50 to 100 individual clones to have an accurate estimation of cytosine methylation, so they favored the direct sequencing approach. However, they also estimated that using the peak height from dye-labeled ddNTP signals was not quantitative enough for an analytic process. So, to still quantify the C signal relative to the T signal from direct sequencing, they used another strategy, which consists in performing the C and T sequencing in separate reactions with the same fluorescent dye to minimize the spectral differences of dyes. And they developed the GENESCAN program to analyze these sequencing results for the relative quantification of C and T signals based on the peak heights of signals.

In the early 2000s, in the context of *The Human Epigenome Project by the Human Epigenome Consortium* (2003), the direct-BSP approach was selected to estimate methylation percentages of high throughput data at a reasonable cost. Hence, it raised the need to have an effective way to compute the percentages based on the sequencing trace data files. The solution was provided by Lewin *et al.* (2004) with the development of an algorithm, called *ESME* for Epigenetic Sequencing Methylation analysis software (Figure 31A) (<https://www.epigenome.org/index.php?page=download>, Lewin *et al.*, 2004). This algorithm aims to compute the CpG methylation levels from the four-dye electropherogram trace file results from the sequencing of PCR products.

As described in the workflow in Figure 31A, the algorithm performs several steps of data processing with quality control steps in between. The first step, called entropy-based clipping, aims to correct base callers artifacts, occurring at the end of the sequencing for example. In this method, they used the area under the trace as the signal intensity value to represent the proportion of the base. The called sequence is aligned with a genomic reference sequence bisulfite converted. Additionally to the common four bases, they used the lower case letter t to represent thymines derived from converted cytosines outside a CpG context which can be matched with either C or T bases (Figure 31B). To remove low-quality results, the flanking regions are clipped to get a remaining inner part containing fewer alignment errors. Then, a step of trace correction aims to correct the potential detection of the mixed C and T signals as distinct positions if there is a small offset between the two peaks. Signals are normalized to compensate for overscaled cytosine traces, as they explain it might be due to the cytosine low frequency compared to other bases. Finally, bisulfite conversion rates can be estimated using the t base positions and methylation percentages are calculated, with the use of the global conversion rate to compensate for incomplete conversions. Based on several experimental

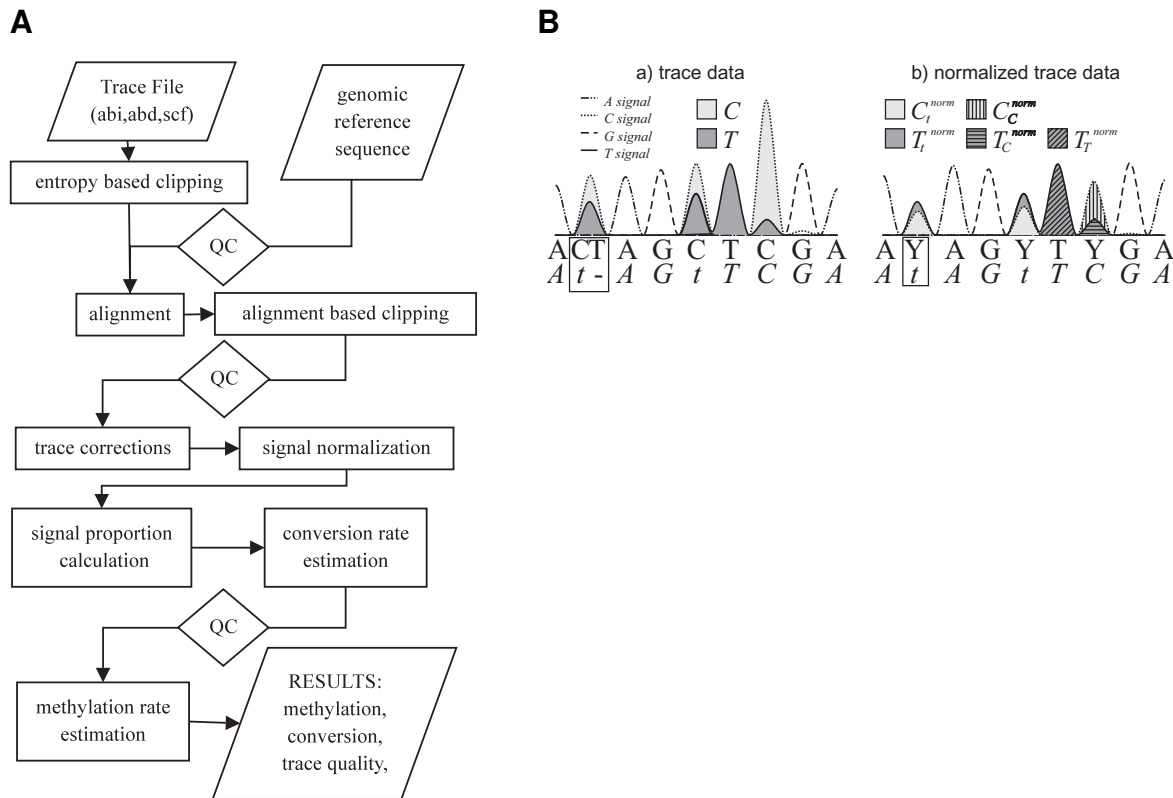


FIGURE 31 THE EPIGENETIC SEQUENCING METHYLATION ANALYSIS SOFTWARE (ESME) TO EVALUATE METHYLATION FROM SEQUENCING TRACE FILES. **A.** Workflow of the *ESME* algorithm. Quality control (QC) steps are performed between each data processing step. The analysis is aborted if the length of the good quality sequence (determined by the clipping) is below the threshold (50 bases by default) or if the bisulfite conversion rates are below the threshold (65% by default). **B.** Normalization step of the *ESME* algorithm. The trace file obtained after direct-BSP is represented (a) before and (b) after the signal normalization. The first line of nucleotides represents the sequence found by the base caller (Y for C/T) and the second line represents the reference sequence, in which t denotes genomic cytosines outside a CpG context, thereby converted into T. The normalization of signals at cytosine positions outside CpG context (t) is needed for evaluation of the bisulfite conversion rate, while the normalization of signals at cytosines position in CpG sites (C) is needed for evaluation of the CpG methylation level. Intensities of the bases are calculated as the area under the trace curve that belongs to the base position. Figures from [Lewin et al. \(2004\)](#).

data, they demonstrated that their algorithm can detect methylation differences with a 20% accuracy and can be applied to high-throughput methylation data.

Currently, it appears that *ESME* is the only tool available for the calculation of methylation levels of CpG based on sequencing trace files for direct-BSP. The majority of recent studies using *ESME* for direct-BSP analysis are studies on patient samples, from a dozen to hundreds of samples, for which the cloning-BSP approach is nearly impossible due to the high cost and time investment. As examples, *ESME* is used in the following recent studies: [Heseding et al. \(2022\)](#); [Ismail et al. \(2020\)](#); [Liu et al. \(2021b\)](#); [Moschny et al. \(2020\)](#); [Schiele et al. \(2021\)](#); [Velásquez et al. \(2021\)](#). Besides providing a solution to analyze direct-BSP data, above all [Lewin et al. \(2004\)](#) demonstrated

the feasibility of a robust methylation levels quantification, with an accuracy estimated at 20%.

The utilization of *ESME* is nonetheless limited due to several issues. The software have been created in 2004 and seems to not have been updated since 2011, and thereby it may not be up-to-date with the evolution and constant improvement of technologies such as sequencing devices, base caller programs, or commercialized bisulfite conversion kits with high conversion efficiencies. For example, the normalization of peaks to compensate for overscaled cytosine signals does not appear to be necessary anymore and might introduce more bias instead of reducing it ([Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems, 2007](#)). Similarly, the compensation for incomplete bisulfite conversion is not relevant nowadays with the variety of ready-to-use bisulfite conversion kits available with efficiencies >99% ([Hernandez *et al.*, 2013](#)). Especially since several studies validated the robustness of results obtained from the manual calculation of methylation percentages using chromatogram peak height ratios without normalization nor bisulfite conversion compensation ([Jiang *et al.*, 2010](#); [Parrish *et al.*, 2012](#)).

Yet, the major problem regarding the use of *ESME* is its accessibility. Available only on Linux operating systems, with a not user-friendly interface, the user needs a high enough proficiency on Linux to install it and run it. Thus, it restricts its use to accustomed users of Linux and might discourage biology researchers ([Akika *et al.*, 2017](#)). Moreover, as it is a 32-bit software, it has to be installed on a 32-bit Linux operating system, yet the widely-used Ubuntu operating system is not available for download in a 32-bit version anymore.

As the purpose of *ESME* was to map CpG methylation levels along the genome, the analysis does not go beyond the calculation of methylation percentages. Indeed, as BSP experiments generally aim to find statistically significant DNA methylation differences between conditions, it requires a visual comparison of results and statistical tests. Thereby, the analysis of results using *ESME* has to be complemented, in most cases, with other tools.

For visualization of methylation data, a web tool has been developed by [Mallona *et al.* \(2014\)](#), *Methylation plotter*, which is implemented as an R shiny app. This tool, accessible on any web browser, provides a dynamic visualization of methylation data by generating a variety of plots, as well as some statistic features ([Figure 32](#)) (http://maplab.imppc.org/methylation_plotter/, [Mallona *et al.*, 2014](#)). Data from both cloning-BSP and direct-BSP methods can be used as inputs as CpG methylation proportions are represented as a grey color gradient from 0 to 1 ([Figure 32A and 32C](#)). Several studies have used *Methylation plotter* to display their direct-BSP results, such as [Forn *et al.* \(2015\)](#); [Gil *et al.* \(2022\)](#); [Ismail *et al.* \(2020\)](#); [Martín *et al.* \(2020\)](#).

Nowadays, even if *ESME* is still used in several studies, the manual calculation of methylation percentages based on peak height was demonstrated to be an accurate method and is the most convenient one to analyze results from direct-BSP experiments ([Jiang *et al.*, 2010](#); [Martisova *et al.*, 2021](#); [Mühlisch *et al.*, 2007](#); [Parrish *et al.*, 2012](#)). This manual analytic process of direct-BSP

data is quite time-consuming, as it depends on the amount of sample and CpG sites per sample, thus, its automatization could increase its efficiency, reduce errors and refine the quality control over sequencing data (Parrish *et al.*, 2012).

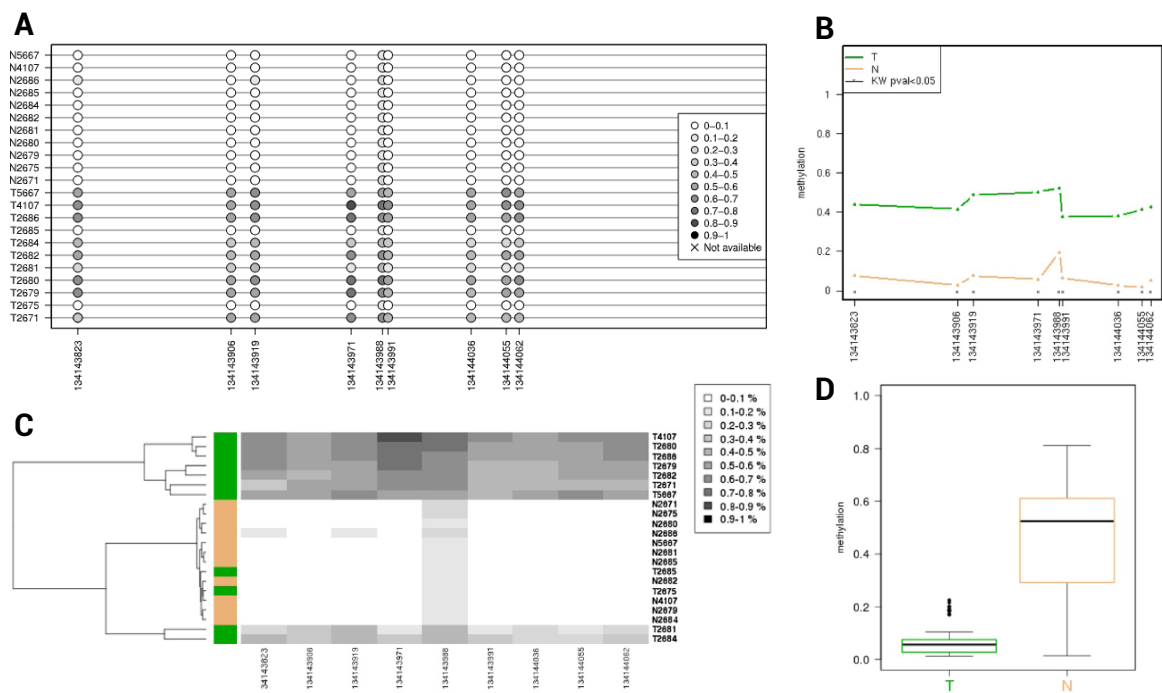


FIGURE 32 VISUALIZATION OF METHYLATION DATA USING THE METHYLATION PLOTTER WEB TOOL. A. Lollipop-like plot, samples are sorted by-group. The normal tissues (N) and tumor tissues (T) present visually different methylation patterns. **B.** Heatmap and its associated dendrogram to display the unsupervised hierarchical clustering of samples, the color corresponds to the different user-provided groups. **C.** Methylation profile plot summarizing the methylation data of groups. An asterisk (*) symbol above the x-axis of CpG sites indicates a statistical difference between groups for the CpG position, according to the non-parametric test Kruskal-Wallis. **D.** Boxplots illustrating the methylation mean and quartiles of each group. Figure adapted from Mallona *et al.* (2014).

KEY POINTS

- ➔ The cloning-BSP approach reveals the methylation status of cytosines, by either a C or a T base, of unique clones. As each clone derives from a unique PCR product, the sequencing of several clones gives an estimation of the CpG methylation proportions in the original DNA pool.
- ➔ Several tools have been developed to analyze cloning-BSP results, using the sequence as input, such as *MethTools 2.0* (Grunau *et al.*, 2000), *BiQ Analyzer* (Bock *et al.*, 2005), *QUMA* (Kumaki *et al.*, 2008) and *BISMA* (Rohde *et al.*, 2010) (Figure 29).
- ➔ In the direct-BSP approach, as a mix of PCR products with different methylation statuses is sequenced, both C and T signals can be measured at cytosine positions. Thus, its analysis relies on the proportionality between the base quantity in the mix of DNA sequenced and its signal intensity on the electropherogram. The methylation percentage can be calculated using the peak height retrieved from the chromatogram: $\text{Methylation percentage} = C/(C+T) \times 100$.
- ➔ The only tool able to analyze direct-BSP results is *ESME* (Epigenetic Sequencing Methylation analysis software) but it suffers from accessibility issues (Figure 31A) (Lewin *et al.*, 2004).
- ➔ The manual calculation of methylation percentage remains the most convenient way to analyze direct-BSP data.
- ➔ Visualization of methylation data from direct-BSP or cloning-BSP can be achieved by using the web-tool *Methylation plotter* (Mallona *et al.*, 2014).

^ [Back to Table of Contents](#)

4.4.4 ARTIFACTS

BISULFITE CONVERSION EFFICIENCY

The most critical artifact to examine during BSP experiments is the conversion efficiency of the bisulfite treatment. When for some unmethylated cytosines the deamination reaction by bisulfite fails, the conversion is described as incomplete. As the methylation percentage determination is based on the relative quantity of either C or T at the CpG site, the incomplete conversion of unmethylated cytosines causes an over-estimation of the C relative quantity and therefore causes methylation levels over-estimation (Genereux *et al.*, 2008; Harrison *et al.*, 1998; Olova *et al.*, 2018; Poucke *et al.*, 2017). In Warnecke *et al.* (2002), the authors reported that, in addition to the commonly described sporadic non-conversion of cytosines, incomplete conversion can also be sequence-specific. They explain that, as the conversion depends on denaturation — or melting — of the DNA molecule, the fragments requiring a higher temperature to melt, such as GC-rich sequences, might be more resistant to conversion.

The bisulfite conversion efficiency can be quantified by calculating the ratio of remaining unconverted C over T at each native non-CpG cytosines position, only if cytosine methylation outside a CpG context is considered rare enough to be negligible, which is the case in mammalian genomes

(Arand *et al.*, 2012; Guo *et al.*, 2014; Laurent *et al.*, 2010; Leontiou *et al.*, 2015). Yet, the most reliable way to evaluate the completeness of conversion is to use unmethylated DNA as a control in bisulfite conversion experiments (Olova *et al.*, 2018). If needed, the conversion efficiency can be improved by altering some parameters like the duration of bisulfite treatment or denaturation temperatures for a given bisulfite concentration (Grunau *et al.*, 2001; Hayatsu *et al.*, 2004). But, for several years now, most of the commercially available kits have been developed and tested to ensure high conversion efficiencies, over 99% in general (Hernandez *et al.*, 2013; Leontiou *et al.*, 2015). Plus, the presence of non-CpGs cytosines converted to thymines in the primer sequences is also a prerequisite to selectively amplify converted DNA molecules and lower this potential artifact (see Appendix 1 “Primer design for bisulfite sequencing PCR” at page 305).

Although the incomplete conversion of DNA is the most described artifact, the inverse artifact also exists and it is called over-conversion of bisulfite. It results from the inappropriate deamination of 5-methylcytosines, thus misinterpreted as unmethylated, which causes methylation levels under-estimation (Genereux *et al.*, 2008; Olova *et al.*, 2018). Yet, the detection of over-conversion by bisulfite is not well reported in the literature.

PCR FIDELITY

The polymerase used for PCR can be a source of bias in the quantification of methylation because of errors introduced during the polymerization (Eckert and Kunkel, 1991; Potapov and Ong, 2017). Indeed, the base substitution at the cytosine position of a CpG site can create a bias in the calculation of the CpG methylation percentage.

For cloning-BSP, if the nucleotide added by the polymerase is a C instead of a T (C>T or G>A in opposite strand polymerization) or a T instead of a C (T>C or A>G in opposite strand polymerization) at the CpG position, it changes the methylation status of the CpG. In Poucke *et al.* (2017), the authors tried to theoretically evaluate the importance of PCR fidelity as a bias in methylation status determination. Based on the theoretical error rate of the *Taq* polymerase, the hypothetical amplification of a 400 bp product containing 40 CpGs for 40 cycles results in an error at 1 CpG per 90 amplicons, which they consider as being 20-fold lesser than the errors introduced by bisulfite incomplete conversion.

For direct-BSP, same as for cloning-BSP, the methylation status can be biased by C>T or T>C (G>A or A>G in opposite strand polymerization) errors as it can shift the methylation percentage in both directions to an over- or under-estimation of methylation. But, the other errors C>A, C>G, T>G, T>A (G>T, G>C, A>C, A>T in opposite strand polymerization) have also a — lesser — impact on the methylation by removing a methylation status information in the total of C+T used in the calculation of the methylation percentage.

Finally, errors introduction in the polymerization can only be reduced by using high fidelity poly-

merase whereas bisulfite conversion-related bias can be strongly reduced by improving experimental conditions. Moreover, PCR fidelity-associated bias seems to be negligible compared to bisulfite conversion bias in the estimation of methylation percentages (Poucke *et al.*, 2017).

BASE MISALIGNMENT IN SEQUENCING RESULT

DNA methylation quantification can also be biased by the misalignment of nucleotides during the Sanger sequencing. The two causes of misaligned base signals are polymerase slippage and mobility difference between unmethylated and methylated fragments.

The presence of long repeats of identical nucleotides, generally of eight or more nucleotides (corresponding to the number of bases in the *Taq* DNA polymerase active site (Eom *et al.*, 1996)), or short motifs repeats (motif up to four bases in general) in the amplicon sequence can induce the polymerase slippage during PCR amplification (Levinson and Gutman, 1987). During extension of a repetitive motif, the slipped-strand mispairing occurs when the polymerase stalls, dissociates from the double-stranded DNA complex and reassociates at another position, one or more repeats ahead or behind the initial point of dissociation. It results in the insertion or deletion of one or more units of the repetitive motif in the newly formed DNA and therefore, the products from PCR can contain amplicons differing from the original template (Sehn, 2015; Shinde *et al.*, 2003). These mutations introduced in some of the amplicons are therefore causing the detection of a mix of base signals at a given position in the Sanger sequencing result (Fazekas *et al.*, 2010; Parrish *et al.*, 2012; *Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems*, 2007). Due to its T redundancy (non-CpG cytosines converted to thymines), bisulfite-treated DNA is prone to polymerase slippage, so it raises the necessity to check for repetitive bases or motifs in the targeted sequence.

The other source of misaligned bases is due to the molecular weight difference between C and T bases. In samples with a mixed methylation state, there is a mix of molecules whose content differs by the presence of C or T at CpG sites. Therefore, the cumulative difference in molecular weight between the DNA templates leads to migration differences during capillary electrophoresis of Sanger sequencing. As differences in C and T accumulates between the different forms of templates, gradually they no longer co-migrate and signals tend to split. For longer fragments, the $n - 1$ base signal of one template can overlap with the n base signal from another template, resulting in an incorrect estimation of methylation levels (Boyd *et al.*, 2006; Boyd *et al.*, 2007; Rakyan *et al.*, 2004). As it is more susceptible to generate bias for longer fragments, limiting the amplicon length is therefore recommended, and limiting the number of CpG sites as well (Rakyan *et al.*, 2004; *Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems*, 2007).

“PCR BIAS”: UNEQUAL AMPLIFICATION OF METHYLATED AND UNMETHYLATED DNA

The methylation-independent PCR (MIP) is based on an equal amplification of the methylated and unmethylated DNA molecules. Indeed, theoretically, as primers do not contain CpG sites, they can anneal to both methylated and unmethylated DNA molecules.

However, it was first described by [Stirzaker *et al.* \(1997\)](#) that this theory is not completely true and that PCR introduces bias. The authors were studying the DNA methylation of a CpG islands located in the retinoblastoma gene (Rb) promoter, in leukocyte DNA from patients. They analyzed the methylation of both top and bottom strands, using MIP primers to amplify bisulfite-treated DNA, and PCR products were directly sequenced or cloned before sequencing. By analyzing both strands, they revealed a preferential amplification of low methylated templates for the top strand sequence and not for the bottom strand sequence in some patient samples. The authors suspected that these results were due to a PCR bias, a difference in amplification efficiencies towards unmethylated DNA, and point out the necessity to test primers on mixed DNA populations with accurate quantification as control.

Temperature melting difference and biased amplification

Then, the same authors in [Warnecke *et al.* \(1997\)](#) first hypothesized that this amplification bias can be due to a higher GC content of methylated DNA, thus impacting the T_m of the DNA molecule, causing the formation of secondary structures and resulting in reduced amplification efficiency, compared to unmethylated T-rich DNA with a lower GC content. This hypothesis then was supported by [Voss *et al.* \(1998\)](#) as the addition of betaine within the PCR reaction, a reagent known to improve amplification of GC-rich DNA forming secondary structures, partially improves the amplification efficiency of methylated DNA ([Aird *et al.*, 2011](#); [Green and Sambrook, 2019c](#); [Henke *et al.*, 1997](#); [Rees *et al.*, 1993](#)).

This preferential amplification towards unmethylated DNA was then described by several other publications ([Guldberg *et al.*, 2002](#); [Moskalev *et al.*, 2011](#); [Rein *et al.*, 1998](#); [Shen *et al.*, 2007](#); [Voss *et al.*, 1998](#); [Warnecke *et al.*, 2002](#); [Wojdacz and Hansen, 2006](#)). Though [Rubatino *et al.* \(2015\)](#) reported an unusual PCR bias towards methylated DNA when amplifying highly methylated regions such as imprinted regions.

DNA melting — or denaturation — is the conversion of double-stranded DNA to single-stranded DNA, also referred to as helix-to-coil transition, which can be achieved by raising the temperature or by denaturing agents. The midpoint temperature at which half of the DNA strands are in the single strand state is called the melting temperature (T_m). The T_m depends on three main factors: the length of the DNA molecule, its base composition, and the ionic strength of the solution (salt concentration). And two main forces maintain the double-stranded conformation of DNA, base pairing: hydrogen bonds between complementary pairs, A:T and G:C, on opposite complementary strands; and base stacking: interactions between neighboring base pairs ([Gotoh, 1983](#); [Vologodskii *et al.*, 1984](#)). Be-

cause there are 2 hydrogen bonds between an A:T pair and 3 hydrogen bond between a G:C pair, they differ in stability, and G:C pairs requires more energy to be broken. Moreover, base stacking interactions are more stable for CG-rich DNA polymers (Vologodskii *et al.*, 1984; Yakovchuk *et al.*, 2006). In other words, the melting of DNA is directly linked to the GC content of the DNA molecule. CG-rich DNA needs a higher temperature to disrupt pair bonds and base stacking interactions for strands to be separated.

In the case of bisulfite-treated DNA, the fully methylated molecule and the unmethylated molecule only differ by their GC content (Figure 33A). The methylated molecule contains C at CpG sites whereas the unmethylated one contains T, and in the opposite strand, G and A respectively. Therefore, the T_m of the methylated DNA is higher than the T_m of the unmethylated DNA (Figure 33B).

This difference can be seen by melting profiles analysis and used to discriminate methylated and unmethylated DNA by methylation-sensitive high resolution melting (MS-HRM). The MS-HRM method can estimate methylation by comparing melting profiles of bisulfite-converted PCR products. Melting profiles are obtained by monitoring the fluorescence changes of dyes intercalated in the double-stranded DNA upon increasing temperatures. As temperature raises, DNA duplexes melt, and fluorescence declines (Wojdacz and Dobrovic, 2007). When analyzing PCR products from bisulfite-treated DNA, the unmethylated DNA melting curve is found different from the methylated DNA melting curve, and in the case of heterogeneous methylated DNA, the melting curve is between those previous two (Figure 33C). By converting the melting curves into melting peaks, as the T_m corresponds to temperature at the maximum peak height, differentially methylated products have therefore different melting peaks (Figure 33D) (Guldborg *et al.*, 2002; Wojdacz and Dobrovic, 2007; Wojdacz and Hansen, 2006).

In consequence, we can extrapolate that the more there are CpG sites in the PCR product, the more GC content difference is high, the more T_m difference between its methylated and unmethylated sequences will be higher. Based on that assumption it can be suggested that reducing as possible the T_m difference by minimizing the amount of CpG sites in PCR products for BSP, to lessen the over-amplification of unmethylated DNA, even though it is not indicated as a guideline in the literature.

Several attempts to modify the melting behavior of DNA during PCR by altering PCR conditions or using additives did not succeed in completely solving this PCR bias (Voss *et al.*, 1998; Warnecke *et al.*, 1997).

Reversal of bias by addition of CpG site in primers

To reverse the over-amplification of unmethylated DNA, Wojdacz and Hansen (2006) submitted the use of primers containing one CpG site in their priming site for MIP. The aim was not to eliminate the bias towards unmethylated DNA but to add another one towards methylated DNA, to counterbalance it, by increasing the selectivity of binding towards methylated DNA molecules. They achieved to reverse the biased amplification with this method, as the expected methylation propor-

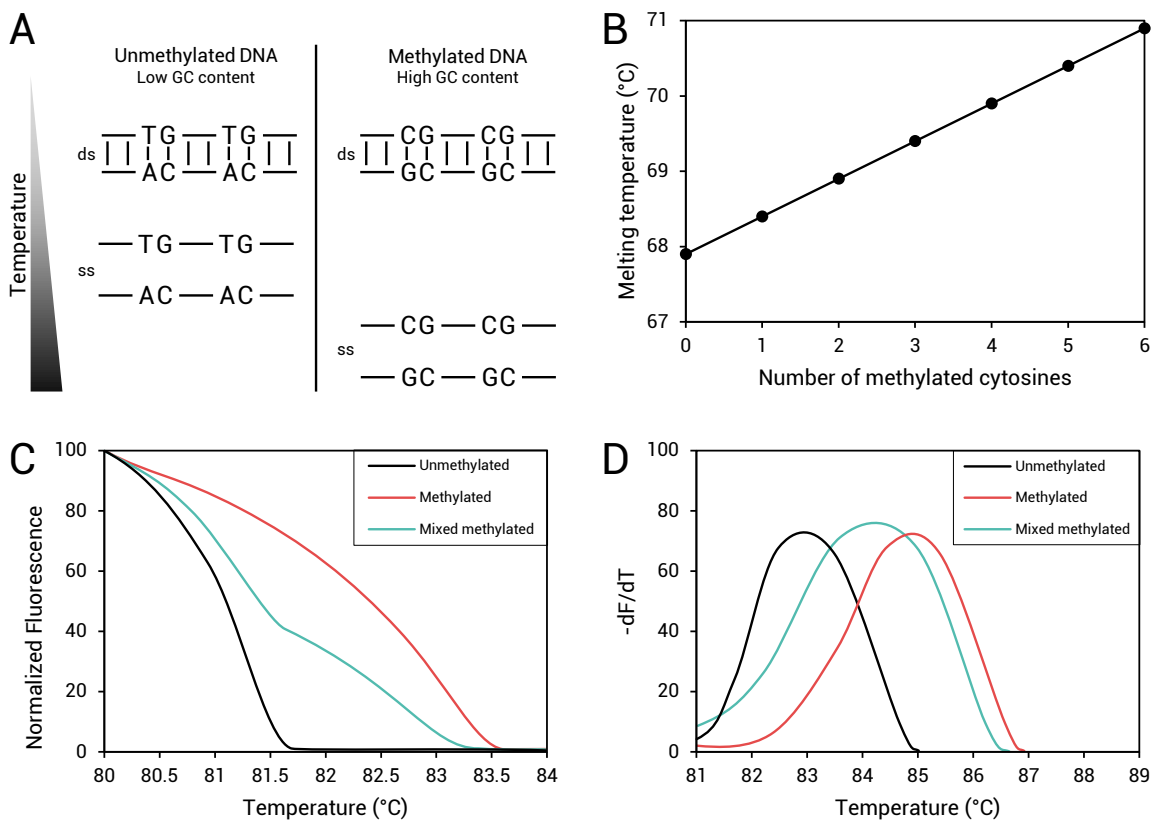


FIGURE 33 MELTING DIFFERENCES BETWEEN UNMETHYLATED AND METHYLATED BISULFITE CONVERTED DNA. **A.** After bisulfite conversion, unmethylated cytosines are transformed in uracil and substituted by thymines during PCR whereas methylated cytosines remain cytosines. Therefore, methylated and unmethylated DNA templates differ in thermal stability due to their different contents of G:C base pairs. ds: double-stranded DNA and ss: single-stranded DNA. **B.** This theoretical graph shows the relationship between T_m and CpG site number. Adapted from [Guldberg et al. \(2002\)](#), it represents T_m values, calculated *in silico*, for an 85 bp domain of a PCR product converted to bisulfite containing, on its 6 CpG sites, from 0 to 6 methylated cytosines. For the ones containing from 1 to 5 methylated cytosines, values represent the averages of the six possible combinations. **C.** This theoretical graph represents melting curves of unmethylated, methylated, and mixed methylated (or heterogeneously methylated) forms of a bisulfite-converted PCR product. **D.** This theoretical graph represents melting peaks of unmethylated, methylated, and mixed methylated (or heterogeneously methylated) forms of a bisulfite-converted PCR product. The melting curves can be converted to melting peaks by plotting the negative derivative of fluorescence over temperature ($-dF/dT$) versus temperature. Figure adapted from [Guldberg et al. \(2002\)](#) and [Wojdacz and Dobrovic \(2007\)](#).

tions were correctly recovered using melting curve assays. Later, the same authors in [Wojdacz et al. \(2009\)](#) compared the use of free-CpG primers and primers including CpG for their ability to proportionally amplify bisulfite-converted DNA in known proportions of methylation, by MS-HRM. They found by adjusting the T_a , that primers with at least one CpG site led to the detection of the expected methylation levels.

The annealing temperature (T_a) is a key parameter during PCR and its optimization is proposed to correct the over-amplification bias. The authors explained that the reversal of this bias

depends on the annealing temperature : at low T_a , primers can bind equally on both methylated or unmethylated templates, thus permitting PCR bias due to differences in efficiencies during the elongation step, but as the T_a is increasing, the primers containing a CpG site can anneal preferably to the methylated template, thus allowing a higher amplification efficiency and the reversal of the bias (Wojdacz *et al.*, 2009).

Reduction of bias by optimizing annealing temperature

Additionally, in Shen *et al.* (2007), bisulfite pyrosequencing was used to assess PCR bias towards amplification of unmethylated DNA and the effect of the T_a on this bias. Most of the primers used contained CpG sites with Y or R bases instead of C/T and G/A bases. By using a mix of unmethylated DNA and methylated DNA in known proportions they were able to confirm a preferential amplification of unmethylated DNA for all of the primer pairs tested, and to overcome or, at least reduce, this bias by increasing the T_a for most of the primer pairs. They could not decipher the relationship between T_a and biased amplification of bisulfite-treated DNA depending on its methylation, yet their theory is that raising the T_a can be sufficient to melt secondary structures of GC-rich DNA and improve its amplification. As the annealing temperature is known to impact the primer affinity to its binding site, it raises the question if the increase of T_a is, rather than melting secondary structures, enhancing the binding affinity of the primer form with the higher T_m (sequence with a C at the CpG site: higher GC content) compared to the primer form with the lower T_m (sequence with a T at CpG site: lower GC content), thus enhancing its annealing on the methylated template, which could explain the bias reduction (Further details in Appendix 1 “Primer design for bisulfite sequencing PCR” at page 305). Yet, the CpG-free primers used by Shen *et al.* (2007) also displayed a biased amplification towards unmethylated DNA which could not be completely solved by increasing the annealing temperature.

Controversy

This method of bias compensation by increasing the selectivity of primers towards methylated DNA is subjected to controversy. In Moskalev *et al.* (2011) the authors found that, for some loci, the presence of a CG dinucleotide in the primer-annealing site can over-compensate the unmethylated DNA over-amplification, leading to the inverse bias towards an over-amplification of methylated DNA. So, instead of trying to optimize PCR to solve bias, they approached the issue by correcting bias afterward on results. Based on curve fitting and using calibration data, they were able to apply correction factors on the results regression curves to correct the methylation percentages estimation.

To find answers to the effect of the presence of CpG sites in the primer annealing sequence on the biased PCR amplification, the authors of Candiloro *et al.* (2017) tested each primer design solution by varying the cytosine of the CpG site and tested different T_a . For the two same primer pairs, they used C-containing primers (methylated), C/T(Y)-containing primers, A/C/G/T(N)-containing primers, inosine(I)-containing primers, mismatch(A)-containing primers, and abasic(-)

containing primers. First, the results from C-containing primers overestimated the DNA methylation for all tested T_a , and this biased amplification towards methylated DNA was enhanced at higher T_a . This finding confirms the over-compensation of PCR bias when the selectivity for methylated DNA is increased by the presence of CpG sites in the primers found by [Moskalev *et al.* \(2011\)](#). Inosine-containing primers showed similar results. For the C/T(Y)-containing primers (G/A(R) in reverse primer) DNA methylation is also overestimated but less than with the previous ones, and the increase of T_a also enhance the bias, which coincides with the results found by [Shen *et al.* \(2007\)](#). The same results were found for the A/C/G/T(N)-containing primers, except for the lowest T_a tested, the DNA methylation estimation was quite accurate. Mismatch-containing primers underestimated DNA methylation and PCR amplification was not efficient. The most accurate estimation of DNA methylation was found using the abasic-containing primers, for only one of the two regions tested as they were not able to amplify the other region using the abasic-containing primers, probably because the abasic site was close to the 3' end of the reverse primer which may interfere with the polymerase priming and activity.

So, in conclusion, it confirmed the over-compensation of PCR bias towards methylated DNA by using primers with CpG sites, whether C-containing primers or C/T(Y)-containing primers, which contradicts the model of primer design proposed by [Wojdacz and Hansen \(2006\)](#). However, most studies agreed that the bias amplification during PCR is region dependant and both models can be relevant depending on the studied region.

Finally, all of the studies emphasized the importance of testing each primer pair to ensure the PCR effectiveness and unbiased results, as well as optimizing the annealing temperature for each primer pair. They recommend performing a gradient of annealing temperature and using a mix of DNA in known proportions of methylation, such as 50:50, as a control to check for the equal amplification of both methylated and unmethylated templates. Therefore, as bisulfite sequencing PCR relies on PCR amplification of heterogeneously methylated DNA, one key step in the experimental process is to carefully check for PCR bias.

KEY POINTS

- ➔ Incomplete conversion of bisulfite can bias the methylation levels estimation. Unmethylated DNA control and calculation of the conversion rate based on non-CpG cytosines are essentials to check for conversion artifacts.
- ➔ Errors in DNA polymerization, to a lesser extent, can also bias the methylation quantification.
- ➔ The misalignment of bases in sequencing results can be due to ① polymerase slippage at repetitive bases or motifs by inducing insertion/deletions mutations and ② gradual migration separation between methylated and unmethylated derived PCR products because of molecular weight difference between C and T bases.
- ➔ A PCR bias causing the unequal amplification of methylated and unmethylated DNA has been widely reported. Most of the time the bias is towards unmethylated DNA due to its lower GC content and thus its lower melting temperature (T_m) (Figure 33).
- ➔ Some attempts to reverse this amplification bias, by including CpG sites in the primer sequences and/or by increasing the annealing temperature (T_a), have been successful for some regions but could also over-compensate the bias towards over-amplification of methylated DNA instead.
- ➔ Incorporating DNA controls with known methylation levels in BSP experiments is essential to check for PCR amplification bias, and optimizing of the T_a for each primer pair can help resolve this bias.

^ [Back to Table of Contents](#)

OBJECTIVES

The cancer stem cells (CSCs) are major contributors to the therapeutic resistance of cancers and an important cause of cancer relapse. Therefore, this project aims to find new solutions to reduce the emergence of CSCs in response to radiotherapy, which leads to the resurgence of the highly resistant CSC population within the tumor. As the cancer non-stem cells (non-CSCs) dedifferentiation into CSCs contributes to the tumor enrichment in CSCs, the inhibition of this phenotypic conversion would radiosensitize tumors and reduce the risk of cancer relapse.

As the acquisition of a stemness phenotype includes the acquisition of self-renewal capacity and pluripotency, it involves changes in the gene expression programs. Hence, we hypothesize that an epigenetic regulation of key genes may regulate the radio-induced dedifferentiation of non-CSCs into CSCs.

In a first part, to get insights into the epigenetic mechanisms involved in the non-CSC-to-CSC conversion, the first objective was to determine the implication of epigenetic modifying enzymes in this process. As modulation of their expression following radiotherapy would indicate their potential participation in this change of phenotype, their expression levels were monitored after radiation exposure. And, to determine the requirement of DNA methyltransferases (DNMTs) enzymes in this process, an inhibition approach was carried out.

Then, genomic regions undergoing methylation from the non-CSC state to the radio-induced CSC state may affect the expression of genes. Therefore, the identification of methylation differences between those phenotypes would allow the discovery of new mechanisms or regulators involved in CSC plasticity. These genes regulated by DNA methylation could be new potential therapeutic targets to specifically inhibit the non-CSC-to-CSC phenotypic switch.

In a second part, the study of DNA methylation levels at specific genomic regions using bisulfite sequencing PCR (BSP) experiments is restrained by an exhaustive analytic process and a lack of efficient tools. Hence, the project aims to provide a new tool capable to handle both direct-BSP and cloning-BSP data, in an automated and accessible way, to help researchers interpret BSP results. Analysis of Bisulfite Sequencing PCR (ABSP) was therefore developed as a ready-to-use solution to facilitate the evaluation of methylation differences in a region of interest.

MATERIAL AND METHODS

1

CELL CULTURE AND CANCER STEM CELLS POPULATION STUDY

☰ CHAPTER CONTENTS

1.1 CELL LINES AND CULTURE CONDITIONS	153
1.2 IRRADIATION	154
1.3 ALDH ACTIVITY STAINING	154
1.4 FLOW CYTOMETRY	154
1.5 FLUORESCENCE-ACTIVATED CELL SORTING (FACS)	155
1.6 SPHERE FORMING CAPACITY	155

[^ Back to Table of Contents](#)

1.1 CELL LINES AND CULTURE CONDITIONS

The SUM159PT triple-negative breast cancer cell line (Asterand) was used and its characteristics are depicted in [Table 5](#). This cell line was cultivated in the culture medium described in [Table 6](#). The Ham's F-12 Nutrient Mixture (F-12), the Non-Essential Amino Acids (NEAA), the HEPES buffer, the penicillin and streptomycin are provided by Gibco; the Fetal Bovine Serum (FBS) by HyClone; the insulin and hydrocortisone by Sigma-Aldrich; and the ZellShield by Biovalley.

SUM159PT	
Supplier	Asterand
Molecular type	Triple negative
Morphology	Epithelial
Tumor type	Anaplastic carcinoma
Sampling origin	Primary tumor

TABLE 5 CHARACTERISTICS OF THE SUM159PT BREAST CANCER CELL LINE.

Cells are maintained in culture in a humid atmosphere at 37 °C containing 5% of CO₂. Before confluence, cells are detached from the Petri dish with trypsin/0,25% EDTA (Gibco), and a portion is reseeded in a new Petri dish. The culture maintenance does not exceed a dozen of seeding.

	SUM159PT
Base medium	F-12
FBS	5%
Insulin	5 µg/mL
Hydrocortison	1 µg/mL
HEPES	10 nM
ZellShield	1%

TABLE 6 COMPONENTS OF THE SUM159PT CELL LINE CULTURE MEDIUM.

1.2 IRRADIATION

Cells are seeded the day before irradiation at an approximated density of 5,500 cells/cm². Cells are irradiated as a monolayer at room temperature in the Oscar Lambret Center of Lille. A 20 MV photon beam is delivered by a Clinac 23EX particle accelerator (Varian) with a dose rate of 2.63 Gy/min. The X- and Y-collimator field is set at 40 cm × 40 cm to irradiate simultaneously up to 16 (4×4) Petri dishes of 10 cm diameter. An 8 Gy dose is delivered at once during a 3,04 minutes lasting run.

1.3 ALDH ACTIVITY STAINING

The high aldehyde dehydrogenase (ALDH) activity of CSCs enables their distinction from the rest of the population using the Aldefluor kit (Stem Cell Technologies) (Ginestier *et al.*, 2007). The Aldefluor kit is used based on the supplier's recommendations. Cells are detached from the monolayer Petri dish and resuspended in the provided Aldefluor assay buffer at a cell density between 5×10^6 and 10×10^6 cells per mL. The reactivated Aldefluor reagent, is added to the cell suspension (5 µL per mL), and cells are incubated for 30 minutes at 37°C into obscurity. Controls used are ① unstained cells incubated in Aldefluor assay buffer only to assess cell autofluorescence and ② stained cells incubated in presence of both Aldefluor reagent and diethylaminobenzaldehyde (DEAB) ALDH inhibitor to define the ALDH^{high} population threshold.

1.4 FLOW CYTOMETRY

The fluorescence of cells stained with the Aldefluor kit is analyzed using either the CyAn ADP flow cytometer (Beckman Coulter) or the CytoFLEX S flow cytometer (Beckman Coulter) at the BioImaging Center Lille (BICeL) platform, located in the IRCL (*Institut pour la Recherche sur le Cancer de Lille*) Institute. The ALDH^{high} population fluorescence intensity threshold is set based on the negative control incubated with an inhibitor of the ALDH activity, the diethylaminobenzaldehyde (DEAB), as it contains only ALDH^{low} cells. ALDH^{high} cells (CSCs) correspond to the cells with a higher fluorescence intensity than the maximum fluorescence intensity of 99.9% of the negative control

population. The cytometry fluorescence data are then analyzed using the FlowJo software (v10.8.1) (BD Biosciences).

1.5 FLUORESCENCE-ACTIVATED CELL SORTING (FACS)

The cells stained using the Aldefluor kit are sorted by FACS using either the BD FACSAria™ III cell sorter (BD Biosciences) or the SH800S cell sorter (Sony) at the BioImaging Center Lille (BICeL) platform, located in the IRCL (*Institut pour la Recherche sur le Cancer de Lille*) Institute. Based on ALDH activity, the ALDH^{low} and ALDH^{high} populations are isolated. The sorted ALDH^{low} population corresponds to 30% of the cell population with the lowest fluorescence intensity. The sorted ALDH^{high} population (CSCs) corresponds to the cells with a higher fluorescence intensity than the maximum fluorescence intensity of 99.9% of the negative control population. During sorting, cells are kept in the Aldefluor assay buffer at 4°C, and sorted cells are retrieved in culture medium at 4°C.

1.6 SPHERE FORMING CAPACITY

The sphere-forming capacity (SFC) assay is a functional assay to estimate the proportion of cells with stem-like capacities. The culture medium used for sphere culture is a DMEM/F-12 (Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12) medium (Gibco) supplemented with 0.4% Bovine Serum Albumin (BSA) (Sigma-Aldrich), 2% B-27 supplement (Gibco), 5 µg/mL of insulin (Sigma-Aldrich), 4 µg/mL of heparin (Sigma-Aldrich), 20 ng/mL of Epidermal Growth Factor (EGF) and Fibroblast Growth Factor (FGF) (Sigma-Aldrich). Cells are seeded in low-attachment surface 96-well plates, from 1024 cells to 1 cell in several wells, with 200 µL of sphere medium per well. The number of formed spheres is measured 8 days later using a phase-contrast microscope. The ratios of the formed sphere over the number of seeded cells in the different dilutions give the estimation of sphere-forming cell proportion. To assess the SFC on several generations, cells are maintained in sphere media in low-attachment surface flasks at a density of 10,000 cells/mL for 10 days per generation. After each generation, cells are dissociated using accutase (Invitrogen) and then reseeded in flasks to get the next generation or in 96-well plates to perform the SFC assay of the current generation.

2

RNA AND PROTEIN LEVELS

CHAPTER CONTENTS

2.1	SIRNA TRANSFECTION	157
2.2	RNA LEVELS	157
2.3	PROTEIN LEVELS	159

[^ Back to Table of Contents](#)

2.1 SIRNA TRANSFECTION

The SUM159PT cells are transfected with siRNAs using the INTERFERin reagent (Polyplus) and the recommended protocol. Cells are seeded one day before to obtain cell confluency of around 50% for transfection. The cell medium is replaced by a fresh medium. The siRNAs are mixed with INTERFERin and cell medium for a quantity of approximately 2.2 pmoles of siRNAs per 200,000 cells. The siRNA sequences used are listed in [Table 7](#). The mix is agitated and incubated for 10 minutes at room temperature before being added to the cell culture. Protein and mRNA relative quantities are measured 48 hours after transfection.

SIRNA NAME	SEQUENCE FORWARD	SEQUENCE REVERSE	REFERENCE
siCtrl	-	-	SR-CL000-05 (Eurogentec)
siDNMT1 #1	5'-CACUGGUUCUGCGCUGGGATT-3'	5'-UCCCAGCGCAGAACCAGUGTT-3'	Fan et al., 2016
siDNMT1 #2	5'-GGAAGUGAAUGGACGUCUATT-3'	5'-UAGACGUCCAUUCACUCCCG-3'	Vispé et al., 2015
siDNMT3B #1	5'-GAUCAAGCUCGCGACUCUCTT-3'	5'-GAGAGUCGCGAGCUUGAUCTT-3'	Fan et al., 2016
siDNMT3B #2	5'-GCUCUUACCUUACCAUCGATT-3'	5'-UCGAUGGUAAGGUAAGAGCTG-3'	Vispé et al., 2015

TABLE 7 LIST OF siRNAs.

2.2 RNA LEVELS

RNA EXTRACTION

The RNA is extracted using the RNeasy kit (Qiagen) following the manufacturer's recommendations. A DNase I treatment during extraction is applied to eliminate genomic DNA. RNA concen-

trations are evaluated by measuring the absorbance at 260 nm by spectrophotometry. RNA samples are stored at -80 °C.

REVERSE TRANSCRIPTION

The cDNA synthesis is performed using the SuperScript III (Invitrogen) reverse transcriptase (RT). As recommended by the manufacturer, RNA and random primers are first incubated for 5 minutes at 65 °C for random primers hybridization, and then the complete reaction mix goes through the reverse transcription steps: 10 minutes at 25 °C, 50 minutes at 50 °C, and 5 minutes at 80 °C. Negative control samples, containing no cDNA, are generated with the same process, except that the reverse transcriptase is not added to the mixture, making the cDNA synthesis impossible.

qPCR

The quantitative polymerase chain reaction (PCR) is performed using the synthesized cDNA, specific primers, and SyBR Green master mix (Qiagen) containing the SyBR Green, the Taq polymerase, deoxynucleotides (dNTP), and MgCl₂. The specific primers used are listed in Table 8, and all of them are synthesized by Eurogentec. The polymerase chain reaction is performed by the CFX96 real-time system (Biorad) thermocycler. It starts with a denaturation step for 10 minutes at 95 °C and is followed by 40 cycles of 3 steps: denaturation (30 seconds at 95 °C), hybridization (30 seconds at the defined hybridization temperature), and elongation (1 minute at 72 °C). The fluorescence measurement is realized at the end of each cycle and the melting curve of the product is generated at the end of the PCR. The optimal hybridization temperature of a primer pair is determined beforehand by evaluating the amplification efficacy at different temperatures. For most of the primer pairs, the optimal hybridization temperature is 59 °C or 60 °C.

TARGET GENE	FORWARD SEQUENCE	REVERSE SEQUENCE
<i>B2M</i>	5'-TCGCGCTACTCTCTCTTT-3'	5'-CAAGTCTGAATGCTCCACTT-3'
<i>RPLP0</i>	5'-GCGACCTGGAAGTCCAATA-3'	5'-TGTCTGCTCCCACAATGAAG-3'
<i>DNMT1</i>	5'-TATCCGAGGAGGGCTACCTGGC-3'	5'-TGGGGCTAGGTGAAGGTTCCAGGC-3'
<i>DNMT3A</i>	5'-TATTGATGAGCGCACAAAGAGAGC-3'	5'-GGGTGTTCCAGGGTAACATTGAG-3'
<i>DNMT3B</i>	5'-TTGAATATGAAGCCCCCAAG-3'	5'-TGATATCCCCTCGTGCTTC-3'
<i>TET1</i>	5'-ATACAATGGGCACCCCTACCG-3'	5'-GGGCTTGGGCTTCTACCAAA-3'
<i>TET2</i>	5'-GCTGACAACTCTACTCGG-3'	5'-CTTCTGGCAAACCTTACATCC-3'
<i>TET3</i>	5'-CCCAAAGAGGAAGAAGTG-3'	5'-GCAGTCAATCGCTATTTC-3'
<i>KDM6A</i>	5'-ATGAATCCTGCAACCAGCCT-3'	5'-TGACTGAGGCCTAATACAGGT-3'
<i>KDM6B</i>	5'-CTCACCGCTATCAGTACCA-3'	5'-GGCAGGATGGATTGACGTT-3'
<i>EZH2</i>	5'-CCCTGACCTCTGTCTTACTTGTGGA-3'	5'-ACGTCAGATGGTGCCAGCAATA-3'

TABLE 8 LIST OF SPECIFIC PRIMERS FOR RT-qPCR.

RESULTS ANALYSIS

Housekeeping genes such as B2M (beta-2 microglobulin) and RPLP0 (Ribosomal Protein Lateral Stalk Subunit P0) are used as reference genes. The expression level fold change calculation is based on the Ct (cycle threshold, cycle at which the fluorescence detected is above the threshold) of each reaction, is normalized using the reference genes, and is relative to the control condition. As each reaction is performed in three replicates, the used Ct value corresponds to the mean of Ct from the three technical replicates. The ΔCt represents the difference between the target gene Ct and the reference gene Ct . The $\Delta\Delta Ct$ represents the difference between the test condition ΔCt and the control condition ΔCt . Therefore, the relative expression fold change corresponds to $2^{-\Delta\Delta Ct}$.

2.3 PROTEIN LEVELS

PROTEIN EXTRACTION

First, for the total protein extraction, cell lysis is performed by the addition of lysis buffer on cells (150 μ L for a 100 mm Petri dish) for 10 minutes at 4°C. The lysis buffer composition is available in Table 9, the NaF, Na₃VO₄, Protein Inhibitor Cocktail (Roche), and phenylmethylsulfonyl fluoride (PMSF) are added right before the lysis. Petri dishes are scraped and the cell lysate is retrieved and centrifuged at 19,000 g for 10 minutes at 4°C. The supernatant containing the protein fraction is stored at -80°C.

COMPONENT	LYSIS BUFFER
HEPES pH 7.5	40 mM
NaCl	120 mM
EDTA pH 8.0	1 mM
NaPPi	10 mM
NaF	50 mM
Na ₃ VO ₄	50 mM
Triton X-100	1%
SDS	0.1%
PMSF	1 mM
glycerol	10%
Protease inhibitor cocktail	1%

TABLE 9 LYSIS BUFFER COMPOSITION FOR TOTAL PROTEIN EXTRACTION.

The cytoplasmic and nuclear protein fractions are isolated using two buffers: a hypotonic one and a hypertonic one. Compositions of the buffers are shown in Table 10. The Protease Inhibitor Cocktail (PIC) (Roche) is added to the required volume of buffer right before the lysis. The cell lysis

starts by adding the hypotonic buffer to adherent cells for 5 minutes at 4°C. Then the samples are placed in microtubes and incubated on a vertical rotating device at 30 rpm (rotation per minute) for 17 minutes at 4°C. After the addition of 2% of non-ionic and non-denaturing detergent Nonidet P-40 (Sigma-Aldrich), the samples are centrifuged at 12,000 g for 5 minutes at 4°C. The supernatant enriched in cytoplasmic proteins is retrieved and stored at -80°C. The pellet is resuspended with the hypertonic buffer and placed once again on a vertical rotating device at 30 rpm for 30 minutes at 4°C. These samples are then centrifuged at 19,000 g for 10 minutes at 4°C and the supernatant enriched in nuclear proteins is retrieved and stored at -80°C.

COMPONENT	HYPOTONIC BUFFER	HYPERTONIC BUFFER
pH	7.8	7.8
HEPES	10 mM	50 mM
KCl	10 mM	50 mM
MgCl ₂	2 mM	-
NaCl	-	300 mM
EDTA	10 mM	10 mM
DTT	3 mM	3 mM
Glycerol	-	10%
Protease inhibitor cocktail	2%	1%

TABLE 10 HYPOTONIC AND HYPERTONIC BUFFERS COMPOSITION FOR CYTOPLASMIC AND NUCLEAR PROTEIN EXTRACTION.

Concentrations in proteins are evaluated by a bicinchoninic acid (BCA) assay (Sigma-Aldrich). After the reaction, the absorbance at 560 nm is measured by spectrophotometry, and concentrations are calculated based on the proportional correlation between absorbance and concentration, using a standard of known concentration.

WESTERN BLOT

Protein extracts are diluted with their corresponding buffer to get equivalent concentrations between each sample. Then, Laemmli buffer, containing SDS and β -mercaptoethanol, is added to the protein extracts for denaturation, for 10 minutes at 95°C under stirring.

Proteins are then separated by gel electrophoresis (SDS-PAGE) in a Mini-PROTEAN Tetra cell (Bio-Rad), using a Bis-Tris 4 to 12% gel (NuPage, Invitrogen) and a MOPS SDS Running Buffer (NuPage, Invitrogen), for 30 minutes at 80 V followed by 1 hour 30 minutes at 120 V, at room temperature. Once separated on the gel, proteins are transferred on a polyvinylidene difluoride (PVDF) membrane (Millipore), using a transfer buffer containing 15% of methanol, for 1 hour and 15 minutes at 105 V, at

4°C. Membranes are saturated by incubation in tris-buffered saline (TBS) buffer supplemented with 0.1% of Tween 20 and 5% of bovine serum albumin (BSA) for 1 hour at room temperature.

TARGET PROTEIN	SUPPLIER	REFERENCE	SPECIE	DILUTION
DNMT1	Abcam	#ab13537	Mouse	1:1000
DNMT3A	Abcam	#ab2850	Rabbit	1:1000
DNMT3B	Abcam	#ab2851	Rabbit	1:1000
Lamin B1	Santa Cruz Biotechnology	#sc-20682	Rabbit	1:500
Mouse IgG	Jackson ImmunoResearch	#115-035-003	Goat	1:5000
Rabbit IgG	Jackson ImmunoResearch	#711-035-152	Donkey	1:5000

TABLE 11 LIST OF ANTIBODIES USED FOR WESTERN BLOT.

Membranes are then incubated with primary antibodies diluted in the saturation buffer overnight at 4°C. The list of used antibodies is displayed in [Table 11](#). After incubation, five washes of the membranes are carried out using TBS buffer 0.1 % Tween 20, for 7 minutes each, under stirring and at room temperature. Then, the membranes are incubated with secondary horseradish peroxidase (HRP)-conjugated antibodies ([Table 11](#)), diluted in TBS buffer 0.1 % Tween 20, for 1 hour and 30 minutes under stirring and at room temperature. Another round of five washes is carried out. Finally, specific proteins bound to the membranes are revealed by the HRP chemoluminescence with the SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) accordingly to the supplier's indications. The numerical images of the membrane are taken using the Fujifilm LAS-4000 Imager. Quantifications of signal intensities of specific bands are computed with the ImageJ software. Actin and lamin B1 proteins are used as loading controls and therefore signal intensities are normalized to these controls.

3

DNA METHYLATION STUDY

☰ CHAPTER CONTENTS

3.1 REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS)	163
3.2 BISULFITE SEQUENCING PCR (BSP)	164

[^ Back to Table of Contents](#)

3.1 REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS)

GENOMIC DNA EXTRACTION

Genomic DNA is extracted using the DNeasy Blood & Tissue kit (Qiagen) and its protocol recommendations. During this experimental protocol, DNA is treated with proteinase K (Qiagen) and ribonuclease A (RNase A) (Sigma-Aldrich). The sample quality is checked by electrophoresis migration using a 0.8% agarose gel. DNA concentrations are obtained by fluorimetry analysis with the Qubit dsDNA HS Assay Kit (Invitrogen) and using the EnSpire Multimode Plate Reader (PerkinElmer). The optimized excitation (478 nm) and emission (526 nm) wavelengths and a standard of known concentrations are used to evaluate the DNA sample concentrations.

RRBS PROCEDURE

The RRBS procedure was carried out by the DNA Methylation profiling Service from Diagenode (Seraing, Belgium). Beforehand, the possible degradation of DNA samples is assessed using the Fragment Analyzer with the DNF-487 Standard Sensitivity or DNF-488 High Sensitivity genomic DNA Analysis Kit (Advanced Analytical). Then, 100 ng of genomic DNA per sample is used to prepare the libraries with the Premium RRBS Kit (Diagenode Cat # C02030033, [Veillard *et al.*, 2016](#)) ([Figure 34](#)).

First, the DNA is digested with the MspI enzyme (C↓CGG, methylation independent) to obtain DNA fragments with CpG sites at each extremity ([Figure 34](#)). Adaptor sequences are ligated to the fragments to allow their size selection (40 to 220 bp) using the Agencourt AMPure XP beads (Beckman Coulter) required to get a high coverage of regions rich in CpG sites. DNA samples are again controlled for concentration and degradation (Qubit dsDNA HS Assay Kit, Thermo Fisher Scientific, and High Sensitivity DNA chip, 2100 Bioanalyzer, Agilent) and pooled. The bisulfite conversion is

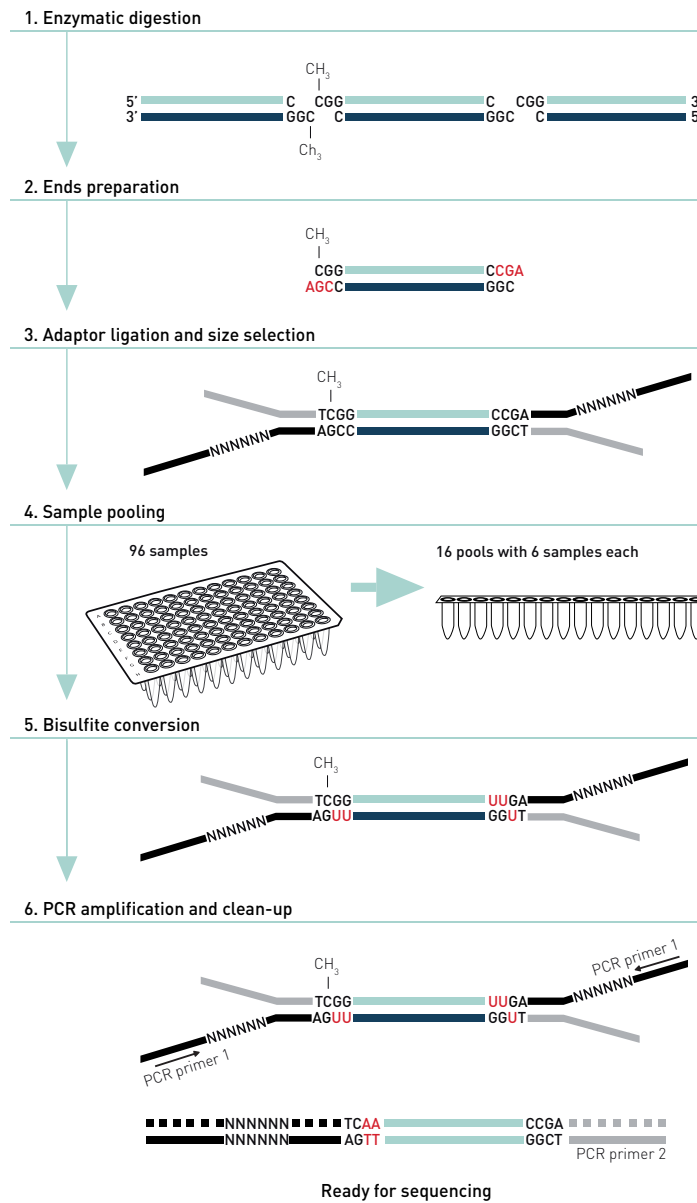


FIGURE 34 REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS) WORKFLOW. Figure from the Premium RRBS kit manual website.

then carried out on the pooled samples to convert unmethylated cytosines into uracils. DNA samples are amplified by PCR and cleaned up to constitute the library ready to be sequenced. So they are finally sequenced by a 50 bp single-read sequencing (SR50) on the HiSeq3000 sequencer (Illumina). Unmethylated and methylated DNA controls serve as controls to check the bisulfite conversion efficiency.

3.2 BISULFITE SEQUENCING PCR (BSP)

BISULFITE CONVERSION

Two DNA samples, high-methylated and low-methylated human genomic DNA (80-8061-HGHM5 and 80-8062-HGUM5 from EpigenDx), are treated with sodium bisulfite. 1.4 µg of each DNA sample is mixed with 0.3 M of NaOH and incubated at 50 °C for 20 min. Then, DNA solutions are treated with a 2.5 M of sodium bisulfite / 125 mM of hydroquinone pH 5.0 solution, at 70 °C for 3 h.

The single-stranded bisulfite converted DNA is then cleaned up using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following the manufacturer's instructions and converted DNA samples are stored at -80 °C before PCR amplification.

PCR

A touchdown **PCR protocol** is used to amplify regions of interest from bisulfite-converted DNA. The composition of each PCR reaction is given in [Table 12](#), and the PCR buffer composition in [Table 13](#). The touchdown PCR protocol is composed of 50 cycles of: 20 s at 95 °C, 30 s at the annealing temperature, and 2 min at 72 °C. The annealing temperature varies from 60 °C for 10 cycles, to 59 °C, 58 °C, 57 °C, and 56 °C for 1 cycle each, and 55 °C for 36 cycles.

COMPONENT	PCR REACTION
PCR buffer 10X	1.5 µL
MgCl ₂ 500 mM	1.2 µL
dNTP 10 mM	0.075 µL
Taq DNA polymerase re-combinant 5 U/µL (Invitrogen)	0.1 µL
Forward primer 5 µM	0.5 µL
Reverse primer 5 µM	0.5 µL
Bisulfite-converted DNA sample	1 µL
H ₂ O	10.125 µL
Total :	15 µL

TABLE 12 PCR REACTION COMPOSITION FOR BISULFITE SEQUENCING PCR.

COMPONENT	PCR BUFFER
Tris pH 8.8	670 mM
(NH ₄) ₂ SO ₄	160 mM
β-mercaptoethanol	100 mM
Bovine serum albumin (NEB)	1 mg/mL

TABLE 13 PCR BUFFER COMPOSITION FOR BISULFITE SEQUENCING PCR.

An upstream promoter region of the CDH1 gene is amplified using the **PCR primers** in **Table 14**. These primers are composed of a sequence-specific to the region of interest and of a standard primer sequence at the 5' end, T3 or BGH Reverse primer sequence (underlined in **Table 14**). The addition of standard sequencing primers to the extremities of the amplicon increases the sequencing efficiency, especially since the beginning is generally not correctly sequenced.

NAME	SEQUENCE	LENGTH	APPLICATION
CDH1 Forward	5'- <u>AATTAACCCTCACTAAAGGG</u> TTTAGTAATTTTAGGTTAGAGGGTTAT-3'	47 bp	PCR
CDH1 Reverse	5'- <u>TAGAAGGCACAGTCGAGG</u> AAACTCACAATACTTTACAATTCC-3'	43 bp	PCR
T3	5'- <u>AATTAACCCTCACTAAAGGG</u> -3'	20 bp	Sequencing
BGH Reverse	5'- <u>TAGAAGGCACAGTCGAGG</u> -3'	18 bp	Sequencing

TABLE 14 PRIMERS USED FOR BISULFITE SEQUENCING PCR.

The **amplified region** is located on the plus strand at coordinates chr16:68771007-68771227 (reference human genome hg19) (221 bp) and covers 17 CpG sites. With the addition of 5' tails — T3 (20 bp) and BGH Reverse (18 bp) — to primers, the total length of the amplicon is 259 bp. Following PCR, amplicon are analyzed by gel electrophoresis migration to validate their correct length.

Amplicon sequence (bisulfite converted sequence with CpG sites considered as methylated, highlighted in yellow, and location of standard primers T3 and BGH Reverse underlined):

5'-AATTAACCCTCACTAAAGGGTTTAGTAATTTTAGGTTAGAGGGTTAT **CGCG**TTTATG**CG**AGGT**CG**GGTGGG**CG**GGT**CG**TTAGTTT**CG**TTTTGGGGAGGGGTT**CGCG**TTGTTGATTGGTTGTGGT**CGGT**AGGTGAATTTT**CG**TTAATTAG**CG**GT**CG**GGGG**CG**GTGTTTT**CG**GGGTTTATTGGTTGTAGTTA**CG**TATTTTTTTTTAGTGG**CGT****CG**GAATTGTAAAGTATTTGTGAGTTT**CG**CCTCGACTGTGCCTTCTA-3'



SEQUENCING

Amplicons are directly sequenced (direct-BSP) in both directions using the standard sequencing primers T3 and BGH Reverse for forward and reverse directions respectively (**Table 14**). Sanger sequencing runs are carried out by the LightRun Sanger sequencing service at Eurofins Genomics, Germany. Three replicates of sequencing runs have been performed to allow statistical analysis. The Applied Biosystems, inc. format (ABIF) (.ab1) sequencing files are used to process the analysis using the ABSP tool.

4

BIOINFORMATICS AND STATISTICS

☰ CHAPTER CONTENTS

4.1 DIFFERENTIALLY METHYLATED REGIONS (DMRs) IDENTIFICATION FROM REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS) DATA	167
4.1.1 TOOLS AND SOFTWARE	167
4.1.2  PACKAGES	168
4.2 DEVELOPMENT OF ANALYSIS OF BISULFITE SEQUENCING PCR (ABSP)	169
4.2.1 TOOLS AND SOFTWARE	169
4.2.2  PACKAGES	169

[^ Back to Table of Contents](#)

4.1 DIFFERENTIALLY METHYLATED REGIONS (DMRs) IDENTIFICATION FROM REDUCED REPRESENTATION BISULFITE SEQUENCING (RRBS) DATA

4.1.1 TOOLS AND SOFTWARE

The **R programming language** is used for the analysis of RRBS data (<https://www.r-project.org/>, R: A Language and Environment for Statistical Computing, version 4.2.0, R core Team, Vienna, Austria, 2022), along with the **RStudio development environment for R** (<https://www.rstudio.com/>, RStudio: Integrated Development for R., version 2022.02.2+485, RStudio Team, PBC, Boston, MA, 2022).

The **IGV (Integrative Genomics Viewer)** (version 2.4.19) is used in the process of analyzing RRBS data to visualize methylation data on the hg19 human genome ([Robinson *et al.*, 2020](#); [Robinson *et al.*, 2017](#); [Robinson *et al.*, 2011](#); [Thorvaldsdóttir *et al.*, 2013](#)).

Pathways enrichment analysis is performed on the **Enrichr enrichment analysis tool** (<https://maayanlab.cloud/Enrichr/>, [Chen *et al.*, 2013](#); [Kuleshov *et al.*, 2016](#); [Xie *et al.*, 2021](#)), and based on the WikiPathways Human 2021 pathways database ([Martens *et al.*, 2020](#); [Slenter *et al.*, 2018](#)). The combined score for each pathway is determined using the p-values from Fisher's exact test and odds ratio.

4.1.2 PACKAGES

For the analysis of RRBS data, different **R packages** are used. The main ones are listed below:

- **BiocManager 1.30.18** - Morgan M (2022). `_BiocManager`: Access the Bioconductor Project Package Repository. R package version 1.30.18, <https://CRAN.R-project.org/package=BiocManager>.
- **biomaRt 2.52.0** - Durinck S, Spellman PT, Birney E, Huber W (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package `biomaRt`. *Nature Protocols* 4, 1184-1191. doi:10.1038/nprot.2009.97 <https://doi.org/10.1038/nprot.2009.97>, <https://www.nature.com/articles/nprot.2009.97>.
- **BSgenome 1.64.0** - Pagès H (2022). `_BSgenome`: Software infrastructure for efficient representation of full genomes and their SNPs. R package version 1.64.0, <https://bioconductor.org/packages/BSgenome>.
- **bsseq 1.32.0** - Hansen KD, Langmead B, Irizarry RA (2012). "BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions." *_Genome Biology_*, *13*(10), R83. doi:10.1186/gb-2012-13-10-r83 <https://doi.org/10.1186/gb-2012-13-10-r83>.
- **ChIPpeakAnno 3.30.1** - Zhu L, Gazin C, Lawson N, Pagès H, Lin S, Lapointe D, Green M (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data." *_BMC Bioinformatics_*, *11*(1), 237. ISSN 1471-2105, doi:10.1186/1471-2105-11-237, <http://www.biomedcentral.com/1471-2105/11/237>.
- **dplyr 1.0.9** - Wickham H, François R, Henry L, Müller K (2022). `_dplyr`: A Grammar of Data Manipulation. R package version 1.0.9, <https://CRAN.R-project.org/package=dplyr>.
- **DT 0.23** - Xie Y, Cheng J, Tan X (2022). `_DT`: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.23, <https://CRAN.R-project.org/package=DT>.
- **ensembldb 2.20.2** - Rainer J, Gatto L, Weichenberger CX (2019). "ensembldb: an R package to create and use Ensembl-based annotation resources." *_Bioinformatics_*. doi:10.1093/bioinformatics/btz031, <https://doi.org/10.1093/bioinformatics/btz031>, <https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz031/5301311>.
- **genomation 1.28.0** - Akalin A, Franke V, Vlahovicek K, Mason C, Schubeler D (2014). "genomation: a toolkit to summarize, annotate and visualize genomic intervals." *_Bioinformatics_*. doi:10.1093/bioinformatics/btu775 <https://doi.org/10.1093/bioinformatics/btu775>, <http://bioinformatics.oxfordjournals.org/content/early/2014/12/04/bioinformatics.btu775.long>.
- **GenomicFeatures 1.48.3** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *_PLoS Computational Biology_*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **GenomicRanges 1.48.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *_PLoS Computational Biology_*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **ggplot2 3.3.6** - Wickham H (2016). `_ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- **Gviz 1.40.1** - Hahne F, Ivanek R (2016). "Statistical Genomics: Methods and Protocols." In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335-351. Springer New York, New York, NY. ISBN 978-1-4939-3578-9, doi:10.1007/978-1-4939-3578-9_16 https://doi.org/10.1007/978-1-4939-3578-9_16, http://dx.doi.org/10.1007/978-1-4939-3578-9_16.
- **knitr 1.39** - Xie Y (2022). `_knitr`: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.39, <https://yihui.org/knitr/>.
- **matrixStats 0.62.0** - Bengtsson H (2022). `_matrixStats`: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.62.0, <https://CRAN.R-project.org/package=matrixStats>.
- **methylKit 1.22.0** - Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012). `_methylKit`: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *_Genome Biology_*, *13*(10), R87. doi:10.1186/gb-2012-13-10-r87 <https://doi.org/10.1186/gb-2012-13-10-r87>, <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-10-r87>.
- **openxlsx 4.2.5** - Schauburger P, Walker A (2021). `_openxlsx`: Read, Write and Edit xlsx Files. R package version 4.2.5, <https://CRAN.R-project.org/package=openxlsx>.
- **png 0.1.7** - Urbanek S (2013). `_png`: Read and write PNG images. R package version 0.1-7, <https://CRAN.R-project.org/package=png>.
- **readr 2.1.2** - Wickham H, Hester J, Bryan J (2022). `_readr`: Read Rectangular Text Data. R package version 2.1.2, <https://CRAN.R-project.org/package=readr>.
- **rlist 0.4.6.2** - Ren K (2021). `_rlist`: A Toolbox for Non-Tabular Data Manipulation. R package version 0.4.6.2, <https://CRAN.R-project.org/package=rlist>.
- **rmarkdown 2.14** - Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2022). `_rmarkdown`: Dynamic Documents for R. R package version 2.14, <https://github.com/rstudio/rmarkdown>.

- **Rsamtools 2.12.0** - Morgan M, Pagès H, Obenchain V, Hayden N (2022). *_Rsamtools*: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import_. R package version 2.12.0, <https://bioconductor.org/packages/Rsamtools>.
- **tidyverse 1.3.1** - Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *_Journal of Open Source Software_*, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

4.2 DEVELOPMENT OF ANALYSIS OF BISULFITE SEQUENCING PCR (ABSP)

4.2.1 TOOLS AND SOFTWARE

The **R programming language** is used to code the ABSP tool (R: A Language and Environment for Statistical Computing, version 4.2.0, R core Team, Vienna, Austria, 2022), along with the **RStudio development environment for R** (RStudio: Integrated Development for R., version 2022.02.2+485, RStudio Team, PBC, Boston, MA, 2022).

The **IGV (Integrative Genomics Viewer)** (version 2.4.19) is used during the development of ABSP to visualize regions on the hg19 human genome, retrieve coordinates and genomic sequences (Robinson *et al.*, 2020; Robinson *et al.*, 2017; Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013).

The BSP primers are designed using the **MethPrimer** program (<https://www.urogene.org/methprimer/>, Li, 2007; Li and Dahiya, 2002), completed with a search for unintended PCR products using the "Primer search, ePCR" from the **BiSearch** tool (<http://bisearch.enzim.hu/>, Arányi and Tusnády, 2007; Arányi *et al.*, 2006; Tusnády *et al.*, 2005).

For the visualization of methylation data, the **Methylation plotter** web tool R source code is used as a reference to build the plot generation functions (http://maplab.imppc.org/methylation_plotter/, Mallona *et al.*, 2014).

4.2.2 PACKAGES

The complete list of used **R packages** is available in the ABSP user guide in Appendix 3 "ABSP user guide" at page 338. The main ones are listed below:

- **arrangements 1.1.9** - Lai R (2020). *_arrangements*: Fast Generators and Iterators for Permutations, Combinations, Integer
- **BiocManager 1.30.17** - Morgan M (2022). *_BiocManager*: Access the Bioconductor Project Package Repository_. R package version 1.30.17, <https://CRAN.R-project.org/package=BiocManager>.
- **Biostrings 2.64.0** - Pagès H, Abooun P, Gentleman R, DebRoy S (2022). *_Biostrings*: Efficient manipulation of biological
- **BSgenome 1.64.0** - Pagès H (2022). *_BSgenome*: Software infrastructure for efficient representation of full genomes and their SNPs_. R package version 1.64.0, <https://bioconductor.org/packages/BSgenome>.
- **compareGroups 4.5.1** - Subirana I, Sanz H, Vila J (2014). "Building Bivariate Tables: The compareGroups Package for R." *_Journal of Statistical Software_*, *57*(12), 1-16. <https://www.jstatsoft.org/v57/i12/>.
- **DiagrammeR 1.0.9** - Iannone R (2022). *_DiagrammeR*: Graph/Network Visu-
- alization_. R package version 1.0.9, <https://CRAN.R-project.org/package=DiagrammeR>.
- **dplyr 1.0.9** - Wickham H, François R, Henry L, Müller K (2022). *_dplyr*: A Grammar of Data Manipulation_. R package version 1.0.9, <https://CRAN.R-project.org/package=dplyr>.
- **formattable 0.2.1** - Ren K, Russell K (2021). *_formattable*: Create 'Formatable' Data Structures_. R package version 0.2.1, <https://CRAN.R-project.org/package=formattable>.
- **GenomeInfoDb 1.32.1** - Arora S, Mor-

- gan M, Carlson M, Pagès H (2022). *_GenomeInfoDb*: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style_. R package version 1.32.1, <https://bioconductor.org/packages/GenomeInfoDb>.
- **GenomicRanges 1.48.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *_PLoS Computational Biology_*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
 - **ggdendro 0.1.23** - de Vries A, Ripley BD (2022). *_ggdendro*: Create Dendrograms and Tree Diagrams Using 'ggplot2'_. R package version 0.1.23, <https://CRAN.R-project.org/package=ggdendro>.
 - **ggplot2 3.3.6** - Wickham H (2016). *_ggplot2*: Elegant Graphics for Data Analysis_. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
 - **ggpubr 0.4.0** - Kassambara A (2020). *_ggpubr*: 'ggplot2' Based Publication Ready Plots_. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.
 - **Gviz 1.40.1** - Hahne F, Ivanek R (2016). "Statistical Genomics: Methods and Protocols." In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335-351. Springer New York, New York, NY. ISBN 978-1-4939-3578-9, doi:10.1007/978-1-4939-3578-9_16 https://doi.org/10.1007/978-1-4939-3578-9_16, http://dx.doi.org/10.1007/978-1-4939-3578-9_16.
 - **htmltools 0.5.2** - Cheng J, Sievert C, Schloerke B, Chang W, Xie Y, Allen J (2021). *_htmltools*: Tools for HTML_. R package version 0.5.2, <https://CRAN.R-project.org/package=htmltools>.
 - **htmlwidgets 1.5.4** - Vaidyanathan R, Xie Y, Allaire J, Cheng J, Sievert C, Russell K (2021). *_htmlwidgets*: HTML Widgets for R_. R package version 1.5.4, <https://CRAN.R-project.org/package=htmlwidgets>.
 - **knitr 1.39** - Xie Y (2022). *_knitr*: A General-Purpose Package for Dynamic Report Generation in R_. R package version 1.39, <https://yihui.org/knitr/>.
 - **openxlsx 4.2.5** - Schauburger P, Walker A (2021). *_openxlsx*: Read, Write and Edit xlsx Files_. R package version 4.2.5, <https://CRAN.R-project.org/package=openxlsx>.
 - **pdftools 3.2.0** - Ooms J (2022). *_pdftools*: Text Extraction, Rendering and Converting of PDF Documents_. R package version 3.2.0, <https://CRAN.R-project.org/package=pdftools>.
 - **plotly 4.10.0** - Sievert C (2020). *_Interactive Web-Based Data Visualization with R, plotly, and shiny_*. Chapman and Hall/CRC. ISBN 9781138331457, <https://plotly-r.com>.
 - **png 0.1.7** - Urbanek S (2013). *_png*: Read and write PNG images_. R package version 0.1-7, <https://CRAN.R-project.org/package=png>.
 - **purrr 0.3.4** - Henry L, Wickham H (2020). *_purrr*: Functional Programming Tools_. R package version 0.3.4, <https://CRAN.R-project.org/package=purrr>.
 - **RColorBrewer 1.1.3** - Neuwirth E (2022). *_RColorBrewer*: ColorBrewer Palettes_. R package version 1.1-3, <https://CRAN.R-project.org/package=RColorBrewer>.
 - **readr 2.1.2** - Wickham H, Hester J, Bryan J (2022). *_readr*: Read Rectangular Text Data_. R package version 2.1.2, <https://CRAN.R-project.org/package=readr>.
 - **renv 0.15.4** - Ushey K (2022). *_renv*: Project Environments_. R package version 0.15.4, <https://CRAN.R-project.org/package=renv>.
 - **rlist 0.4.6.2** - Ren K (2021). *_rlist*: A Toolbox for Non-Tabular Data Manipulation_. R package version 0.4.6.2, <https://CRAN.R-project.org/package=rlist>.
 - **rmarkdown 2.14** - Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2022). *_rmarkdown*: Dynamic Documents for R_. R package version 2.14, <https://github.com/rstudio/rmarkdown>.
 - **Rmisc 1.5.1** - Hope RM (2022). *_Rmisc*: Ryan Miscellaneous_. R package version 1.5.1, <https://CRAN.R-project.org/package=Rmisc>.
 - **rstatix 0.7.0** - Kassambara A (2021). *_rstatix*: Pipe-Friendly Framework for Basic Statistical Tests_. R package version 0.7.0, <https://CRAN.R-project.org/package=rstatix>.
 - **sangeranalyseR 1.6.1** - Chao K, Barton K, Palmer S, Lanfear R (2021). "sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R." *_Genome Biology and Evolution_*. doi:10.1093/gbe/evab028 <https://doi.org/10.1093/gbe/evab028>.
 - **sangerseqR 1.32.0** - Hill JT, Demarest BL, Bisgrove BW, Su Y, Smith M, Yost HJ (2014). "Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products." *_Developmental Dynamics_*. doi:10.1002/dvdy.24183. <https://doi.org/10.1002/dvdy.24183>.
 - **shiny 1.7.1** - Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2021). *_shiny*: Web Application Framework for R_. R package version 1.7.1, <https://CRAN.R-project.org/package=shiny>.
 - **shinythemes 1.2.0** - Chang W (2021). *_shinythemes*: Themes for Shiny_. R package version 1.2.0, <https://CRAN.R-project.org/package=shinythemes>.
 - **webshot 0.5.3** - Chang W (2022). *_webshot*: Take Screenshots of Web Pages_. R package version 0.5.3, <https://CRAN.R-project.org/package=webshot>.

RESULTS

1

DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION

☰ CHAPTER CONTENTS

1.1 INVOLVEMENT OF EPIGENETIC-ASSOCIATED ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION	173
1.1.1 EXPRESSION OF EPIGENETIC-ASSOCIATED ENZYMES AFTER IRRADIATION	173
1.1.2 ROLE OF THE DNMT ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION	176
📌 KEY POINTS	180
1.2 GLOBAL ANALYSIS OF DNA METHYLATION CHANGES DURING RADIO-INDUCED DEDIFFERENTIATION	180
1.2.1 STRATEGY AND SAMPLE PREPARATION	180
1.2.2 FIRST ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA	181
1.2.3 SECOND ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA	189
1.2.4 IDENTIFIED DIFFERENTIALLY METHYLATED REGIONS (DMRs)	197
📌 KEY POINTS	204

[^ Back to Table of Contents](#)

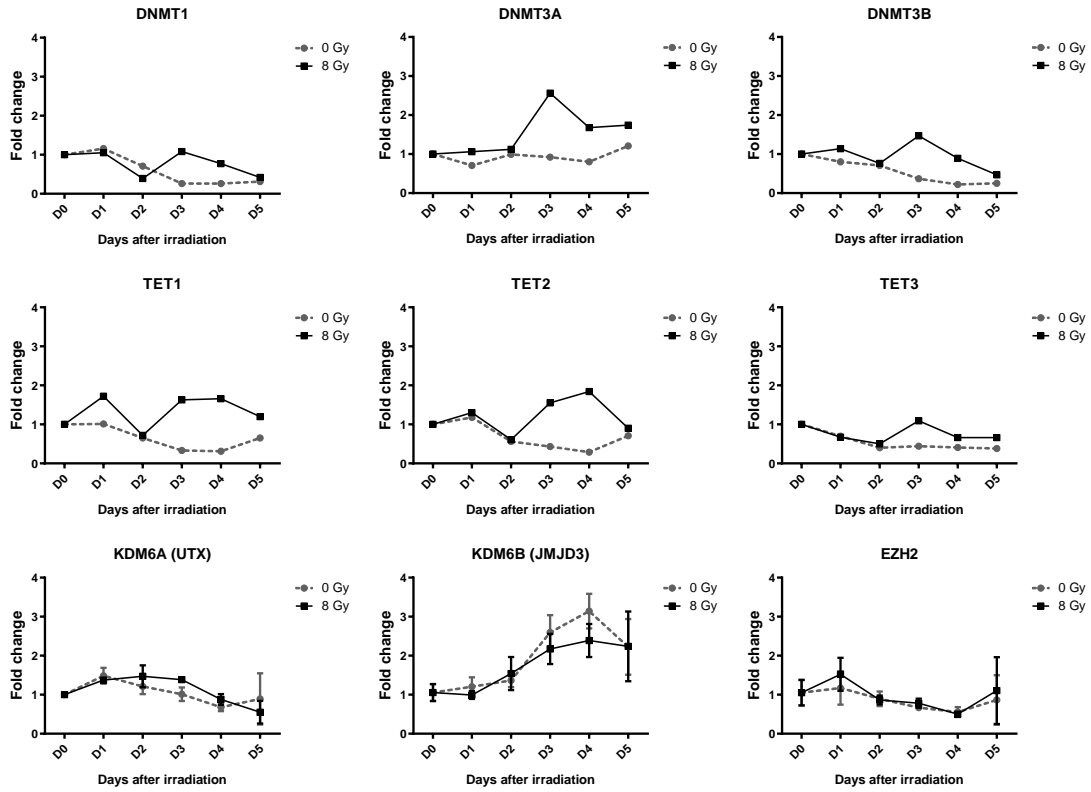
1.1 INVOLVEMENT OF EPIGENETIC-ASSOCIATED ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION

Upon and following radiotherapy treatment, some breast non-CSCs are able to dedifferentiate into CSCs as they acquire stem cell properties (Arnold *et al.*, 2020b; Bidan *et al.*, 2019; Lagadec *et al.*, 2012). As previously described, epigenetic mechanisms are involved in this process as they can modify gene transcription programs toward stemness-related genes expression (French and Pauklin, 2021; Zagorac *et al.*, 2016). Hence, we hypothesize that, to remove or establish new epigenetic marks, the activity of epigenetic-associated enzymes must be required and contributes to the non-CSC-to-CSC phenotypic switch after ionizing radiation exposure.

1.1.1 EXPRESSION OF EPIGENETIC-ASSOCIATED ENZYMES AFTER IRRADIATION

The first element that can indicate the participation of epigenetic modifying enzymes throughout the radio-induced dedifferentiation is to evaluate their expression following radiotherapy. Indeed,

A



B

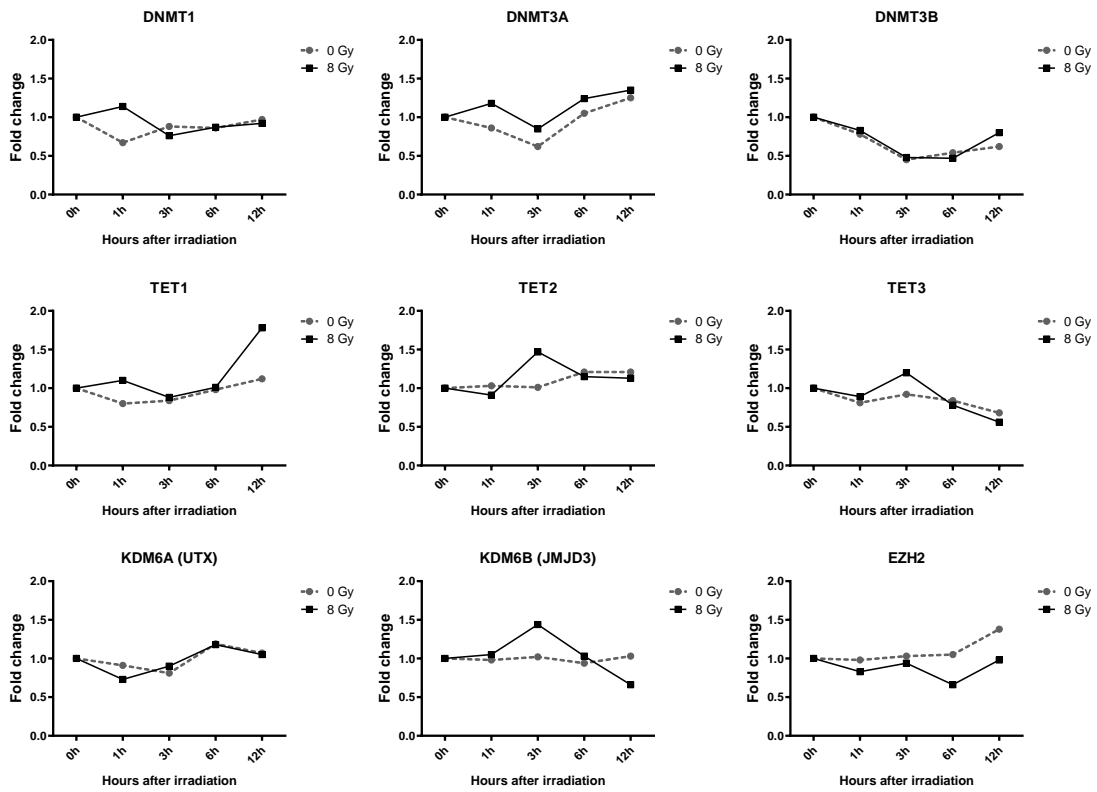


FIGURE 35 EXPRESSION LEVELS OF THE EPIGENETIC-ASSOCIATED ENZYMES FOLLOWING IRRADIATION. **A.** SUM159PT cells were irradiated at 8 Gy and the mRNA levels of DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, and EZH2 were measured every day up to 5 days post-irradiation. For DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, n=1 and for KDM6A, KDM6B, EZH2, n=3. Error bars represent means \pm standard error of the mean. **B.** ALDH^{low} SUM159PT cells were isolated by FACS reseed and irradiated at 8 Gy 24 hours later. mRNA levels of DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, and EZH2 were measured 1 hour, 3 hours, 6 hours, and 12 hours after irradiation (n=1).

an increase in their expression could be correlated with enhanced recruitment to apply epigenetic modifications, to alter the transcription of gene programs.

To cover different types of epigenetic marks such as DNA methylation and histones PTM, either towards the transcription activation or silencing, I chose nine key enzymes to study: DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, and EZH2. Indeed, the DNA methyltransferases (DNMTs) are responsible for the maintenance and establishment of DNA methylation marks on cytosines while TET enzymes are involved in active DNA demethylation (Ginno *et al.*, 2020; Kim and Costello, 2017). EZH2 is the PRC2 subunit that catalyzes the methylation of the lysine 27 on histone H3 to set H3K27me3 repressive marks, while KDM6A (also known as UTX) and KDM6B (also known as JMJD3) are two lysine demethylases (KDMs) that remove H3K27me3 marks (Pediconi *et al.*, 2019).

To assess the effect of radiotherapy treatments on these epigenetic enzymatic activities, I measured their expression through time (every 24 hours) in SUM159PT breast cancer cells after irradiation at 8 Gy (Figure 35). The effect of an 8 Gy radiation treatment applied to this cell line was already assessed in the laboratory and is known to induce an enrichment in CSCs, with a peak at 5 days post-irradiation (Bidan *et al.*, 2019; Lagadec *et al.*, 2012). Therefore, the first assay, presented in Figure 35A, aimed to estimate mRNA levels once a day during 5 days post-irradiation. For DNMT3A, a small increase in mRNA levels can be observed but only at day 3 (D3). A slight increase of TET1 and TET2 expression (1.5-1.8 fold compared to D0) can be observed 3 days (D3) and 4 days (D4) after irradiation at 8 Gy, which is 4 to 5 times higher than in the unirradiated cells. It's worth to notice that these differences at D3 and D4 in expression are transitory and not maintained up to 5 days after irradiation (D5). Concerning KDM6A, KDM6B, and EZH2, no significant difference was observed between the irradiated and unirradiated cells among three independent experiments. Taken together, as no significant differences in epigenetic enzyme expression were detected between irradiated and unirradiated cells, no conclusion on their involvement on dedifferentiation can be drawn from this experimental approach.

As modifications of expression following irradiation can occur earlier on the first day, I chose to assess their mRNA levels in the first 12 hours post-irradiation. Additionally, to observe the effect of the phenotypic switch from non-CSCs to CSCs instead of CSC pool repopulation by self-renewal and

symmetric division upregulation, especially since they might require different expression changes, the population was depleted in CSCs before irradiation. The ALDH^{low} population was isolated by FACS the day before the radiation treatment and mRNAs were extracted at 0h, 1h, 3h, 6h and 12h post-irradiation (Figure 35B). No evidence of expression changes was observed after the first experiment, and as the dedifferentiation occurs within the non-CSC population in non-synchronized manner along the 5 days after irradiation, we decided to not push further this short term study.

Furthermore, two possible explanations can support the absence of expression differences observed after irradiation: ① as the radio-induced dedifferentiation rate is low, no more than 5% of CSCs are induced in general, the evaluation of changes in the global population does not allow the observation of modifications associated with these rare events, and ② the modulation of their activity, possibly involved in the dedifferentiation, could depend on co-activators rather than on their expression. Indeed, the action of these enzymes may involve other regulators to modulate their recruitment at specific genomic loci to enable the phenotypic switch toward a stem state.

In conclusion, the approach of analyzing expression changes in the global population does not allow the identification of fine expression regulation of DNMTs, TETs, KDM6A, KDM6B, and EZH2 enzymes in a specific subpopulation undergoing phenotypic changes after radiation exposure.

1.1.2 ROLE OF THE DNMT ENZYMES IN RADIO-INDUCED DEDIFFERENTIATION

The previous results indicate that the expression of DNMTs enzymes is not upregulated following radiotherapy, but their activity may be involved in the phenotypic conversion. To know if DNMTs enzymes are required for the dedifferentiation of non-CSCs into CSCs, their inhibition by small interfering RNA (siRNA) during the radio-induced dedifferentiation was carried out.

First, two siRNAs targeting DNMT1 and DNMT3B were tested, named siDNMT1 #1 and #2 and siDNMT3B #1 and #2. Both siRNAs were able to inhibit the two enzymes at RNA and protein levels 48 hours after transfection (Figure 36). As siDNMT1 #2 and siDNMT3B #2 seem to be the most efficient ones, they were chosen for the following experiments. Four different siRNAs targeting DNMT3A were tested but failed to inhibit the enzyme, so the study of DNMT3A has not been possible.

To evaluate the importance of DNMT1 and DNMT3B epigenetic modifying activities in the reacquisition of stem cell properties, they were inhibited by the previously validated siRNAs before the radio-induction of dedifferentiation in the non-CSC population. As described in Figure 37A, the ALDH^{low} cells (non-CSCs) are isolated by FACS and treated with the siRNAs 18 hours after being reseeded. Then, 6 hours after siRNA transfection, cells are irradiated at 8 Gy. Induction of dedifferentiation toward a stem cell phenotype was evaluated 5 days later by quantifying ALDH⁺ cells (CSCs) by flow cytometry measurement and sphere-forming capacity (SFC) tests. A negative control is used to define the ALDH⁺ population for each sample, in which stained cells are incubated with DEAB, an

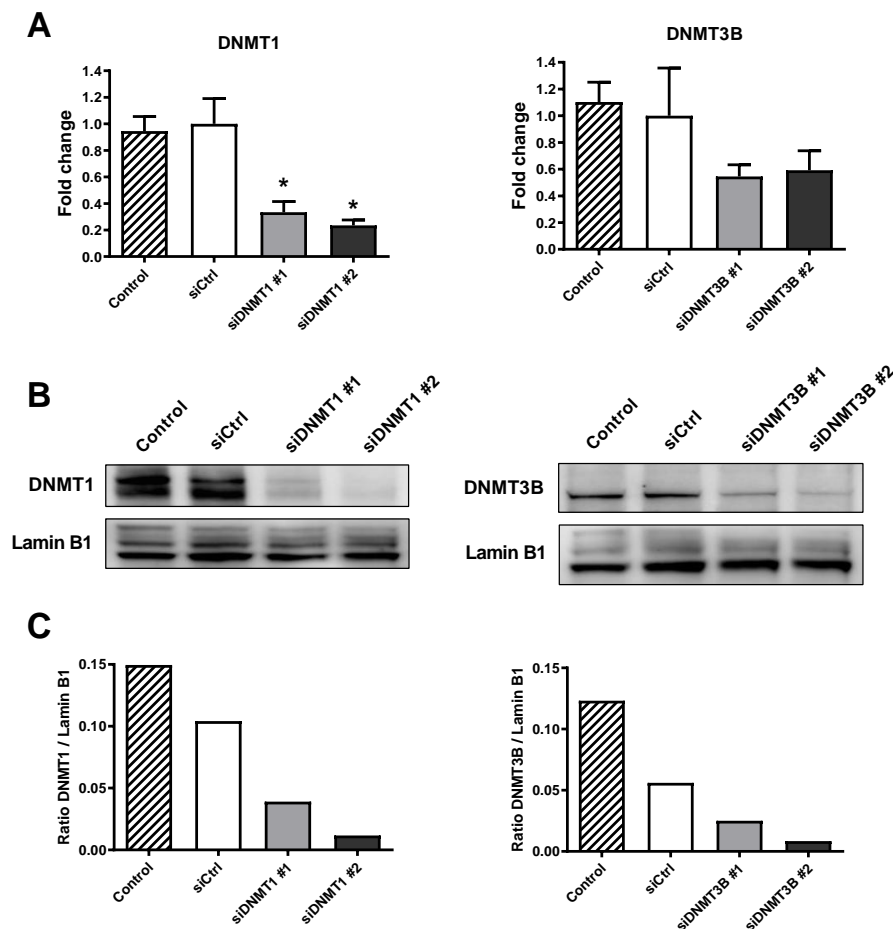


FIGURE 36 INHIBITION OF DNMT1 AND DNMT3B BY siRNA INTERFERENCE. **A.** The mRNA expression levels of DNMT1 and DNMT3B in SUM159PT cells were measured 48 hours after siRNA transfection ($n=3$). The control condition corresponds to untransfected cells. Assessment of the inhibition is confirmed by comparing siDNMT1/3B conditions with the respective siCtrl condition. * Student's T-test p -value ≤ 0.05 . Error bars represent means \pm standard error of the mean. **B.** The DNMT1 and DNMT3B inhibition is confirmed at the protein level using nuclear extracts collected 48 hours after siRNA transfection. The lamin B1 protein is used as a loading control. The same results were obtained in three independent experiments. **C.** Quantification of band intensities from the above blots results. Ratios of the DNMT1/3B band intensities over the lamin B1 band intensities are represented as a bar chart.

ALDH inhibitor (Figure 37A and B).

First, dot plots of analyzed cells from one of the four experiments are displayed in Figure 37B. The cell population is distributed in terms of their Aldefluor intensity, reporting for the ALDH activity, and size (side scatter). In the control untransfected population, a shift toward higher Aldefluor intensities is visible in the irradiated population compared to the unirradiated one, corresponding to an increase of the CSC population in response to radiations. However, in all the transfected conditions, siCtrl, siDNMT1, and siDNMT3B, no shift in the irradiated population is observed, meaning that in these cells the irradiation did not induce an increased dedifferentiation of non-CSCs.

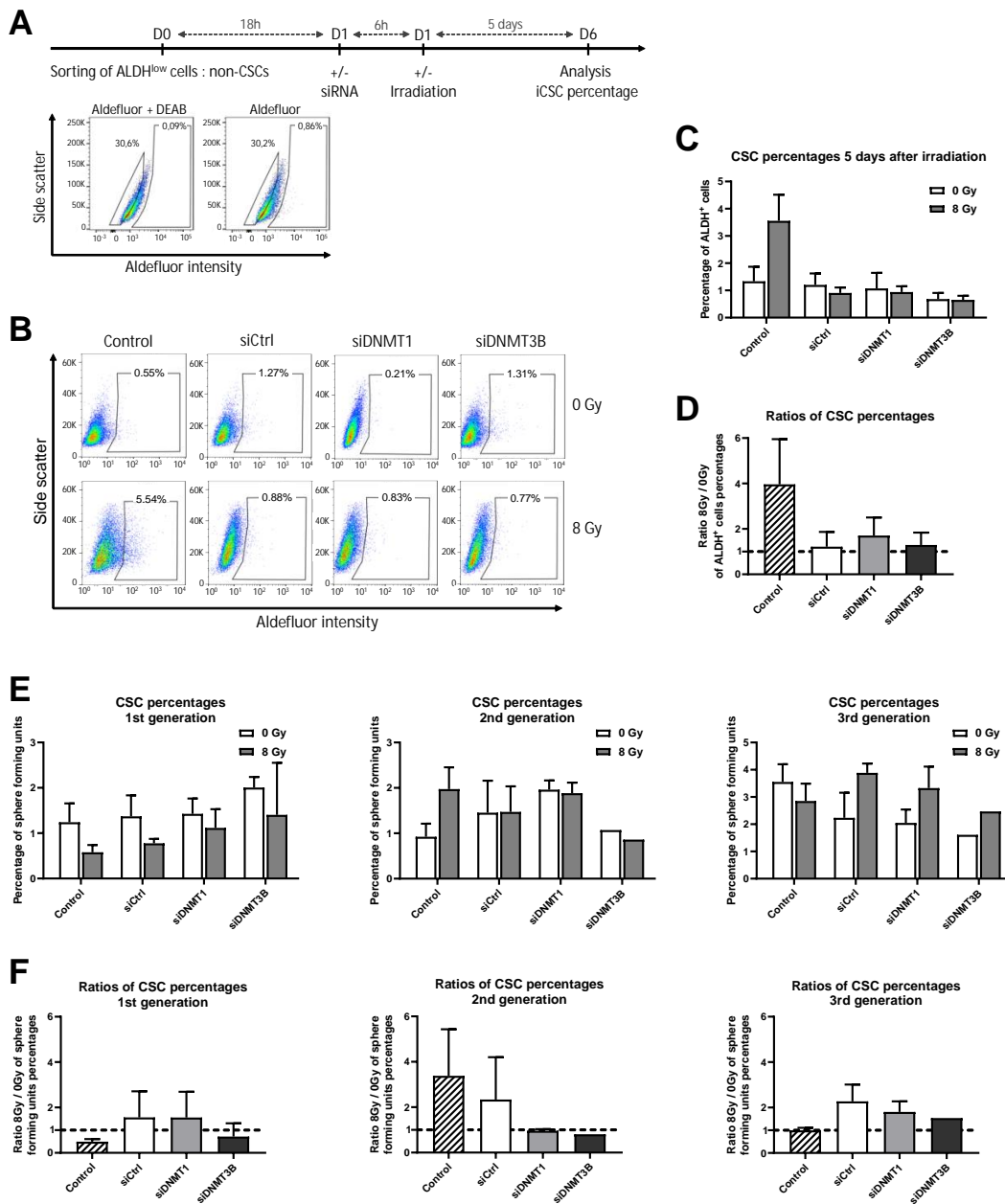


FIGURE 37 EFFECT OF DNMT1 AND DNMT3B INHIBITION ON THE DEDIFFERENTIATION OF NON-CSCS INTO CSCS. **A.** Experimental protocol to assess the dedifferentiation of non-CSCs. The SUM159PT ALDH^{low} cells (non-CSCs) are sorted by FACS, treated the next day with siRNAs 18 hours later and with radiations 6 hours later. The generation of induced ALDH⁺ (iCSCs) is evaluated 5 days after irradiation. **B.** Cytometry analysis of non-CSCs SUM159PT cells stained with Aldefluor after siRNA transfection and irradiation. The gate and its associated percentage represent the ALDH⁺ (CSC/iCSC) population. **C.** Percentage of ALDH⁺ cells induced by irradiation in the control (n=4). **D.** Ratios of irradiated cells CSC percentages over unirradiated cells CSC percentages from the above cytometry data (n=4). **E.** Percentages of sphere-forming unit estimated by sphere-forming capacity (SFC) assay, after one generation (n=4), two generations (n=3, n=1 for siDNMT3B), and three generations (n=3, n=1 for siDNMT3B). **F.** Ratios of irradiated cells sphere forming unit percentages over unirradiated cells sphere forming unit percentages from the above SFC assay data. Error bars represent means ± standard error of the mean.

Means of four independent experiments are exposed in **Figure 37C**, in which no significant differences using the Student's T-test was observed. In the control, the non-CSC population restores after 5 days an average of $1.34\% \pm 0.53$ of CSCs without induction, and after radiation exposure, it regenerates an average of $3.56\% \pm 0.96$ of CSCs. These data indicate an increased generation of CSCs induced by irradiation even if the difference is not statistically significant ($p=0.104$) due to important variability among the four replicates. Nonetheless, while we can see a drastic reduction of the CSC induction from $3.56\% \pm 0.96$ in the irradiated control cells to $0.94\% \pm 0.21$ and $0.65\% \pm 0.15$ in irradiated siDNMT1 and siDNMT3B conditions respectively, no induction has been also observed in the siCtrl condition ($0.91\% \pm 0.20$) after irradiation. It indicates that the suppression of the radio-induced dedifferentiation is due to the siRNA transfection protocol and is not linked to the inhibition of DNMT1 or DNMT3B. In **Figure 37D** the difference in CSC percentages between unirradiated cells and irradiated cells is displayed as averages of the 8 Gy over 0 Gy ratios per replicate, which also highlights the increase of CSCs in control cells but not in transfected cells.

CSC proportions were also estimated by a functional SFC test with three generations of sphere formation (**Figure 37E and F**). Cells are maintained in sphere media to allow the formation of spheres for 10 days before being dissociated to form a new generation of spheres. At each generation, the number of spheres is measured and reveals the proportion of cells able to self-renew. Consistent with previous lab results, an important mortality in the irradiated conditions causes a reduced sphere formation at the first generation, but it stabilizes at the second and third ones. However, a slight increase in sphere forming unit proportions after irradiation in control cells is observed at the second generation but not in the third one. And altogether, in every transfected conditions at every generation, no significant difference in CSC proportions is observed, which is consistent with the previous cytometry results, suggesting that the transfection protocol prevents the induction of CSCs.

Furthermore, the combination of FACS sorting followed by siRNA transfection and irradiation caused an important loss of cell viability in cells. Even though different transfection protocols were tested (INTERFERin, Lipofectamine, and nanoparticles), all of the transfected cells had considerably lower viability than the control untransfected cells. Thus, the siRNA transfection, even with a control siRNA without any specific target, is sufficient to drastically reduce the cell viability, which can explain differences between control and siCtrl conditions in terms of CSC radio-induced regeneration.

As the transfection itself seemed to abolish the radio-induced increase of CSCs, probably due to cell viability issues, the contribution of DNMT enzymes in the non-CSC-to-CSC conversion could not have been assessed by the siRNA inhibition approach used, and would require an alternative approach.

KEY POINTS

- The expression analysis of the global population after radiotherapy treatment did not reveal significant expression changes of the epigenetic modifying enzymes DNMTs, TETs, KDM6A, KDM6B, and EZH2.
- The approach consisting in a global assessment of expression differences at a few time points after irradiation could not reveal transitory expression changes related to rare, asynchronous, and spread over time dedifferentiation events.
- As the radio-induced increase of CSC observed in control cells is not reproduced in transfected cells, including cells transfected with a control siRNA, it means that the transfection itself impacts the radio-induced regeneration of CSCs.
- Hence, the effect of the siRNA-mediated DNMT1 and DNMT3B inhibition on the radio-induced dedifferentiation could not have been assessed. An alternative approach is required to further evaluate the contribution of these enzymes in the phenotypic conversion from non-CSCs to CSCs.

[^ Back to Table of Contents](#)

1.2 GLOBAL ANALYSIS OF DNA METHYLATION CHANGES DURING RADIO-INDUCED DEDIFFERENTIATION

The reduced representation bisulfite sequencing (RRBS) approach was chosen to assess methylation marks profiles before and after irradiation in the subpopulations in question, to determine methylation modifications occurring during the radio-induced dedifferentiation of non-CSCs into CSCs. The RRBS method allows for the computation of cytosine methylation percentages on a reduced fraction of the genome, enriched in CpG sites and representative of the genome methylome. The aim is to identify differentially methylated regions that might be involved in the acquisition of a stem-like phenotype.

1.2.1 STRATEGY AND SAMPLE PREPARATION

CSC and **non-CSC** subpopulations of SUM159PT breast cancer cells were collected after a first FACS sorting based on the ALDH activity (Figure 38). The non-CSC population is reseeded and irradiated at 8 Gy the next day to induce the dedifferentiation of cells into CSCs. Five days later, the population is sorted once again to isolate the **induced CSC** population by irradiation, named **iCSC**, and the **irradiated non-CSC** population, named **inon-CSC**. iCSC cells underwent a radio-induced dedifferentiation event while the inon-CSC ones did not after radiation exposure. As seen previously, the SUM159PT cell line is relatively rich in CSCs, with a basal CSC percentage of around 1 to 2%, and the irradiation treatment of non-CSCs generates 2.5 to 5% of iCSCs after 5 days (Figure 37A and C, Figure 38, and data not shown).

DNA from the four collected subpopulations, **CSC**, **non-CSC**, **iCSC** and **inon-CSC**, in two replicates each, was extracted and sent to Diagenode for the RRBS analysis, consisting in MspI

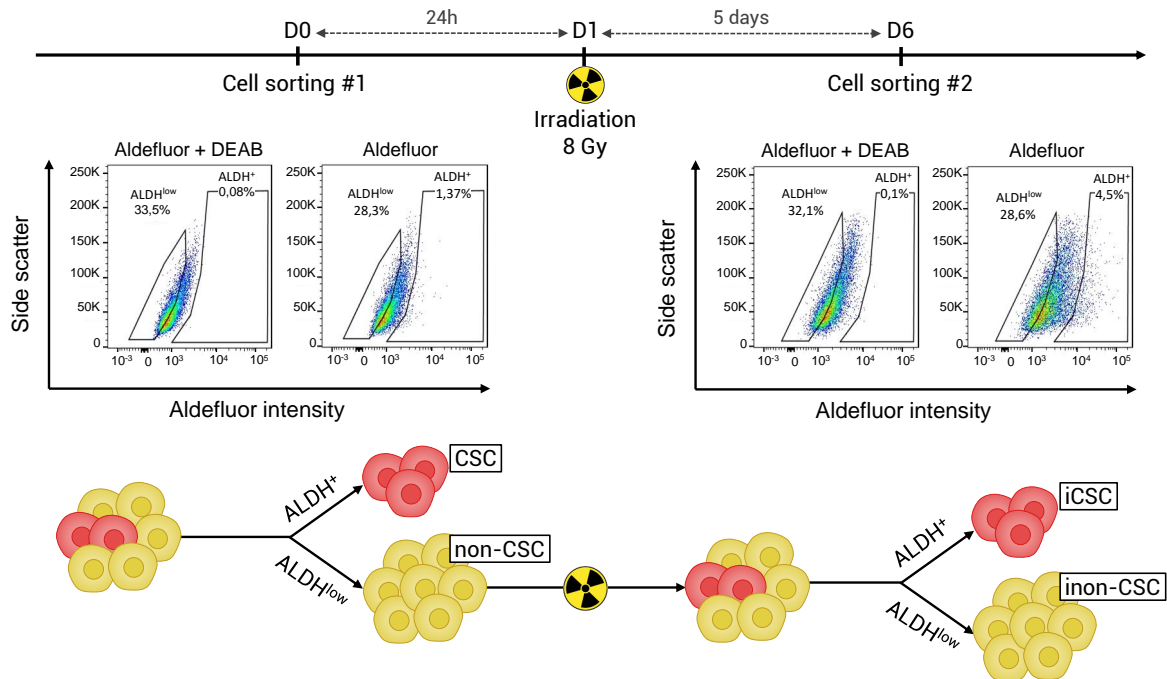


FIGURE 38 EXPERIMENTAL RADIO-INDUCTION OF DEDIFFERENTIATION AND SUBPOPULATION ISOLATION FOR RRBS ANALYSIS. Two rounds of cell sorting based on ALDH activity are performed to isolate four SUM159PT subpopulations: CSC, non-CSC, iCSC, and inon-CSC populations. The CSC and non-CSC samples are collected after a first sorting. The isolated non-CSC population is reseeded and then irradiated at 8 Gy 24 hours later. The irradiated non-CSC population is sorted 5 days post-irradiation to isolate induced CSCs (iCSCs) and irradiated non-CSCs (inon-CSCs).

enzymatic digestion (C↓CGG, methylation independent), bisulfite treatment, and next-generation sequencing (NGS). Reads aligned on the genome resulting from the sequencing were then analyzed to identify methylation differences between those subpopulations.

1.2.2 FIRST ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA

The analysis of RRBS data aims to identify differentially methylated regions (DMRs) between the different analyzed subpopulations to highlight potential key methylation changes contributing to the phenotypic switch into CSC after radiotherapy.

A first analysis of RRBS results was performed with the help of the bioinformatic service platform Bilille (PLBS - UMS 2014 - US 41, Lille, France) for DMRs identification. This analysis was carried out using the R software and relies mainly on the `methy1Kit` R package developed specifically to analyze RRBS data (Akalin *et al.*, 2012).

This analysis consists in first subdividing the genome into 1,000 bp tiles (i.e. chr1:1-1,000; chr1:1,001-2,000; chr1:2,001-3,000;...) to calculate the differential of methylation based on methylation percentages of CpG sites within each tile.

DMRs are generated by comparing methylation levels of subpopulations two by two. For a

1,000 bp region to be considered differentially methylated, the **methylation difference must be > 25%** (absolute value, > 25% for hypermethylated ones, and < -25% for hypomethylated ones) between those two populations with a **q-value < 0.01** for statistical significance. The method used to calculate the differential methylation and its associated significance relies on a Bayesian hierarchical model, based on the beta-binomial distribution, which is provided by the `methyKit` R package, but developed originally within the `DSS` R package (Feng *et al.*, 2022; Feng *et al.*, 2014; Feng and Wu, 2019). Indeed, as methylation proportions are comprised between 0 and 1 (0% and 100% methylated) their distribution follows a beta distribution among replicates, with a binomial distribution captured by the sequencing within each sample (Feng *et al.*, 2014).

Once methylation differences between subpopulations have been found, the goal is to find the ones potentially implicated in the dedifferentiation of non-CSCs into iCSCs (Figure 39). The direct comparison of non-CSC vs iCSC populations is not sufficient. Indeed, it includes methylation modifications due to irradiation which are not necessarily related to the stem cell phenotype.

- **DMRs between non-CSC vs CSC constitute the base pool of DMRs** used for the analysis (Figure 39 and Figure 40A, in yellow). These DMRs corresponds to CSC-specific methylation differences that may be involved in the stemness phenotype. The regulation of these regions by DNA methylation could be linked to pluripotency or differentiated features.
- **DMRs between CSC vs iCSC are excluded** (Figure 39 and Figure 40A, in pink). The exclusion of differences between CSC and iCSC populations (the two CSC populations) allows to keep from the base DMRs pool the ones that are methylated back in the same way after dedifferentiation, meaning that in the CSC to non-CSC to iCSC phenotypic path, these regions underwent methylation changes both from CSC to non-CSC and from non-CSC to iCSC, their methylation status is restored in iCSCs similarly to CSCs.
- **DMRs between non-CSC vs inon-CSC are excluded** (Figure 39 and Figure 40A, in blue). The methylation differences between non-CSC and inon-CSC populations (the two non-CSC populations) were not sufficient to induce the phenotypic switch into iCSC, they underwent methylation changes in response to irradiation that did not lead to the acquisition of an iCSC state. Therefore, only the regions with consistent methylation from non-CSC to inon-CSC populations are kept.

From the four subpopulations, the three pairwise comparisons generates DMRs:

- **non-CSC vs CSC:** 1,340 DMRs (Figure 39 and Figure 40A, in yellow)
- **CSC vs iCSC:** 969 DMRs (Figure 39 and Figure 40A, in pink)
- **non-CSC vs inon-CSC:** 833 DMRs (Figure 39 and Figure 40A, in blue)

As displayed in Figure 40A, from the 1,340 DMRs obtained comparing non-CSC vs CSC, 159 also different between CSC and iCSC are excluded, 65 also different between non-CSC and inon-CSC are excluded, and 5 different in the three pairwise comparisons are excluded as well. Hence, a list of **1,112 sorted DMRs potentially involved in the radio-induced dedifferentiation** is retrieved,

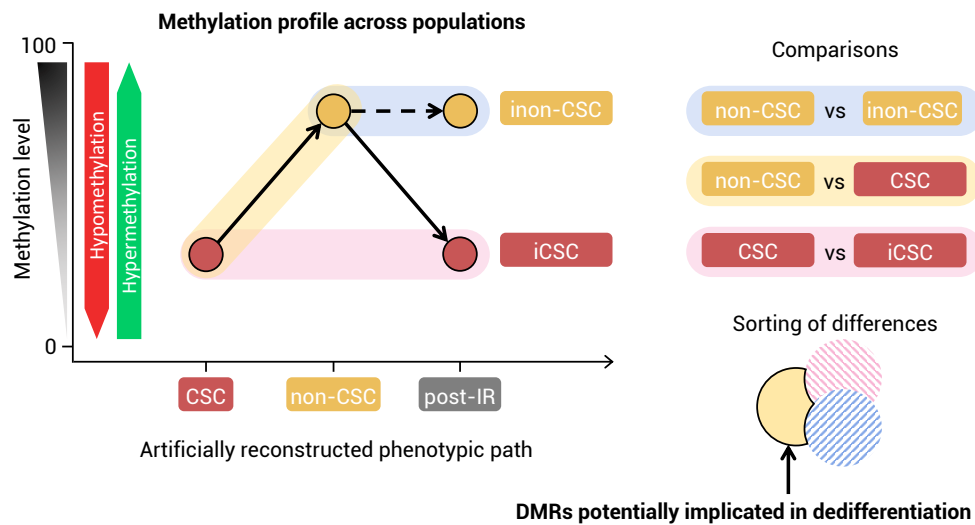


FIGURE 39 SCHEMATIC REPRESENTATION OF THE METHYLATION PROFILE OF A DMR POTENTIALLY INVOLVED IN DEDIFFERENTIATION. Hypothetical methylation profile of a region that could be involved in radio-induced dedifferentiation through the DNA methylation regulation of its associated gene. This DMR is hypomethylated in CSCs compared to non-CSCs, and no differences are found between the two CSC states and the two non-CSC states. The reverse pattern can be applied to hypermethylated DMRs in CSCs compared to non-CSCs.

containing 681 hypomethylated DMRs and 431 hypermethylated DMRs (in CSCs compared to non-CSCs).

The next objective is to find the most relevant DMRs for a more precise region-specific validation of methylation percentages. To refine this DMR list, I retrieved the closest gene of each DMR, within a maximal distance of 10 kb. From the 1,112 DMRs, 596 are within the 10 kb range of 558 unique genes, 332 hypomethylated ones, and 264 hypermethylated ones. The pathway analysis of this list of genes revealed several genes from the **Ephrin receptor signaling** (e.g. EFNB1, EPHA2, EPHB2), **TGF- β signaling** (e.g. BMP2, BMP4, BMP7, MAP2K2, MAPK13, RUNX3, TGFB3) and **STAT3 signaling** (e.g. IGF1R, IL17RB, IL4R, MAP2K2, MAP3K9, MAPK13, NTRK3), known to be involved in the stem cell phenotype and resistance to radiotherapy of breast cancer cells (Bhatia *et al.*, 2018; Huang *et al.*, 2017; Lucero *et al.*, 2020; Yadav and Shankar, 2019; Yang *et al.*, 2013).

Additionally, DMRs positions were cross-referenced with the CpG island (CGI) database (UCSC), and from the list of 1,112 DMRs, 63 are found overlapping a CGI. From the previously cited genes, only the DMR associated with the BMP4 gene overlaps a CGI.

As these DMRs were found as differentially methylated between non-CSC and CSC populations, there is a potential correlation between these identified methylation changes and their closest gene expression in non-CSC vs CSC populations.

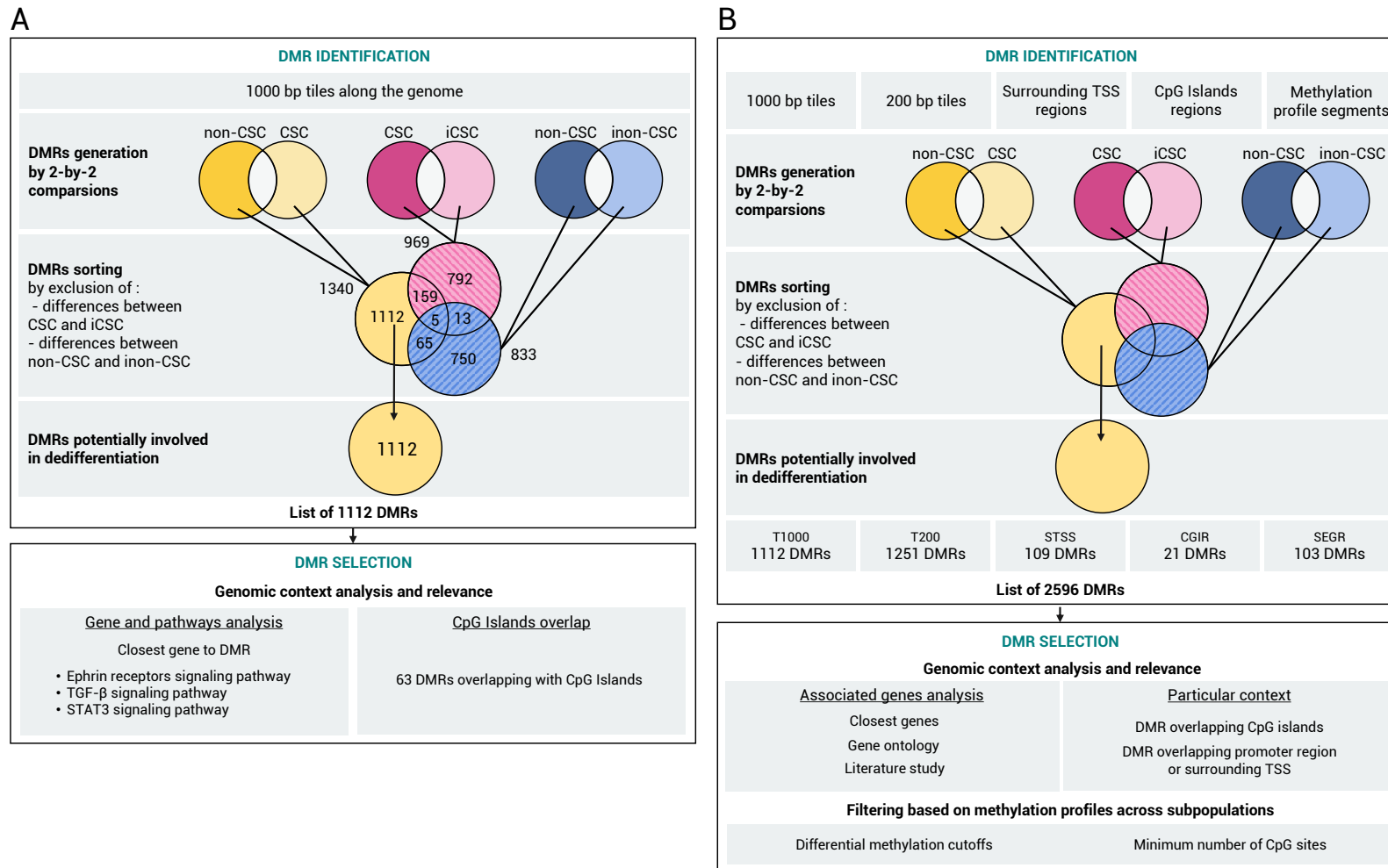


FIGURE 40 RRBS ANALYSES DIFFERENCES BETWEEN THE FIRST AND THE SECOND ONE. **A.** The first analysis of RRBS data was based solely on the 1,000 bp tiles and limited filtering of the identified DMRs. **B.** The second analysis realized afterward incorporates the new subset of genomic regions in addition to 1,000 bp tiles: 200 bp tiles, surrounding TSS regions (500 bp upstream and downstream TSS, 1,001 bp long in total), CGI regions, and methylation profile segments. Additionally, further DMR filtering enables the identification of the most relevant ones.

To test the functional role of these methylation differences, 5 DMRs were selected, close to these 5 genes: **EPHB2**, **RUNX3**, **MAPK13**, **TGFB3**, and **BMP2**, based on the previous pathway analysis associated to literature study.

Among them, the RUNX3 promoter is known to be frequently hypermethylated in mammary tumors (Lau *et al.*, 2006; Liu *et al.*, 2018b; Lu *et al.*, 2016; Song *et al.*, 2016). RUNX3 is regulated by the TGF- β , its gene code for a tumor suppressor, and its expression is correlated to stem cell differentiation and reduction of stem cell properties (Kim *et al.*, 2019a; Kulkarni *et al.*, 2018; Lee *et al.*, 2010; Liu *et al.*, 2020a; Nishina *et al.*, 2011; Wang *et al.*, 2017e). The DMR found around its TSS containing one CpG site covered in all samples, is hypomethylated in CSCs compared to non-CSCs, which is inconsistent with the literature, but could be due to the poor coverage of the region.

Methylation percentages of CpG sites contained in these 5 DMRs are represented along the genome in Figure 41A and their methylation profiles across subpopulations in Figure 41B. The methylation profiles of DMR#21–EPHB2 and DMR#22–RUNX3 show a higher similarity in methylation levels between non-CSC and iCSC populations than between non-CSC and inon-CSC populations, while the methylation profiles of DMR#324–MAPK13, DMR#742–TGFB3 and DMR#967–BMP2 show a no difference between inon-CSC and iCSC populations (high standard deviations) despite having a significant methylation difference between non-CSC and CSC populations. Thereby, their profile does not exactly match the desired pattern of DMRs undergoing methylation changes during radio-induced dedifferentiation as displayed in Figure 39, and the analysis and selection of relevant methylation profiles should be taken into consideration further in the RRBS data analysis. However, as they all present an important methylation difference between non-CSC and CSC, this methylation difference could be correlated with gene expression changes.

To assess the correlation between methylation differences and expression, the CSC and non-CSC populations were sorted based on their ALDH activity and the expression of these 5 genes was analyzed by RT-qPCR. For those 5 DMRs, no significant expression differences were found between the two subpopulations (Figure 41C). EPHB2 and MAPK13 expressions seem to be up-regulated in CSCs compared to non-CSCs, which is inconsistent with hypermethylated DMRs found inside these two genes (EPHB2 gene locus: chr1:23,037,330-23,241,823 and DMR#21 locus: chr1:23,176,001-23,177,000; MAPK13 gene locus: chr6:36,098,260-36,112,301 and DMR#324 locus: chr6:36,101,001-36,102,000). Hence, there is no correlation between the expression of these genes and the identified differential methylation, so the validation of these DMRs was not pursued.

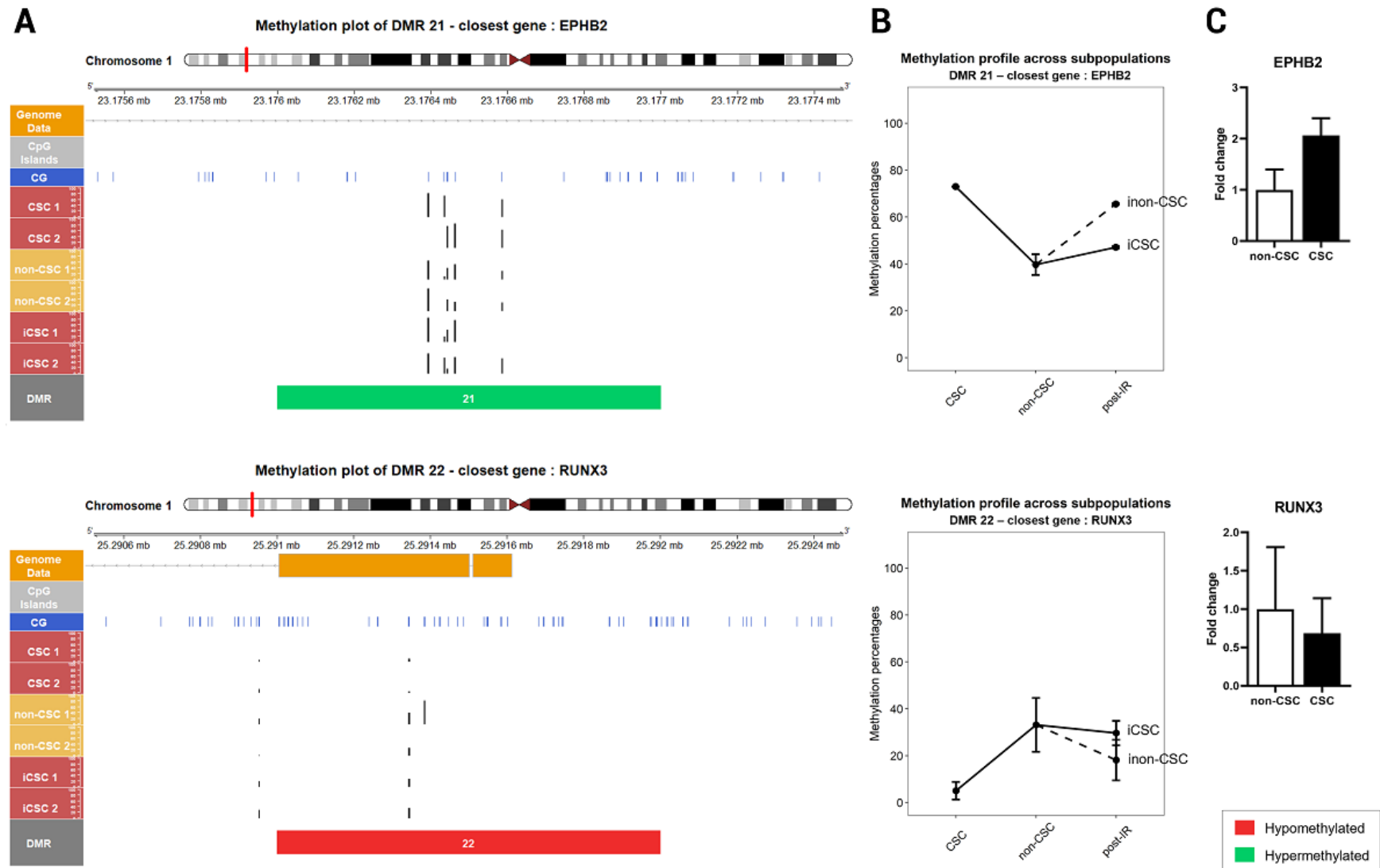


FIGURE 41 Caption in following pages.

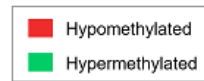
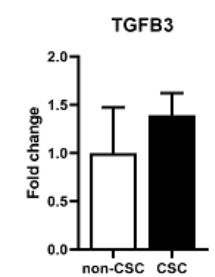
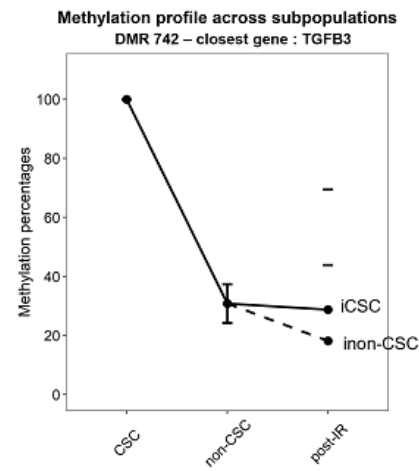
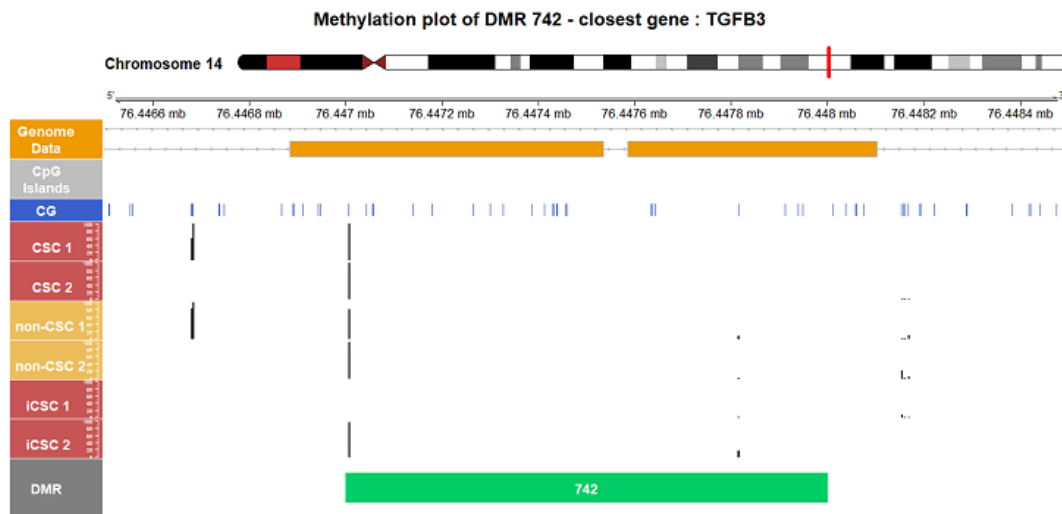
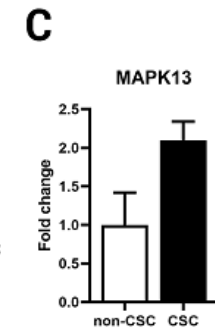
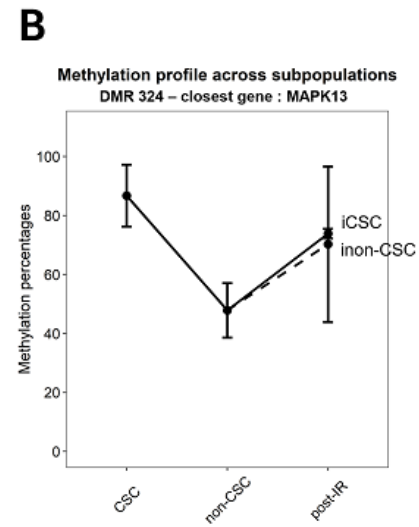
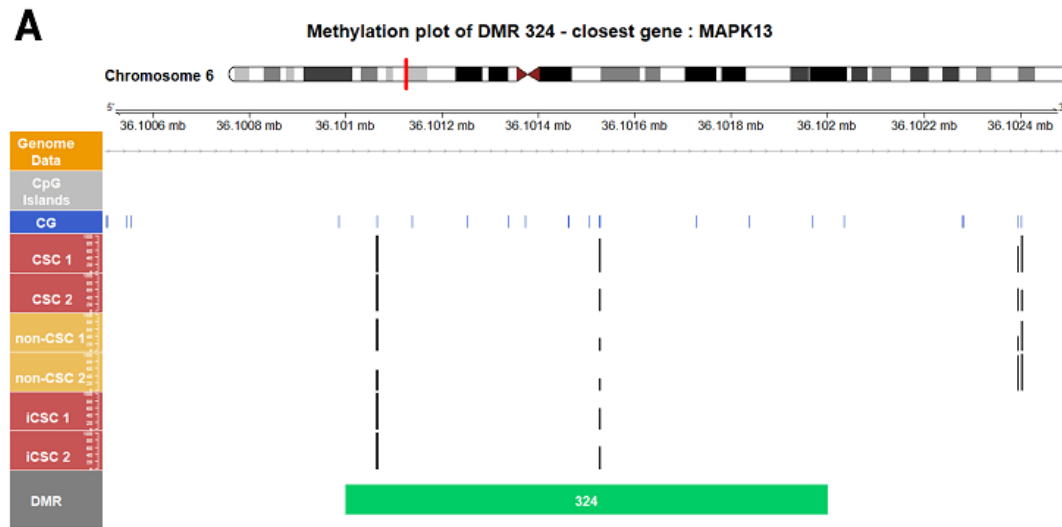


FIGURE 41 *Caption next page.*

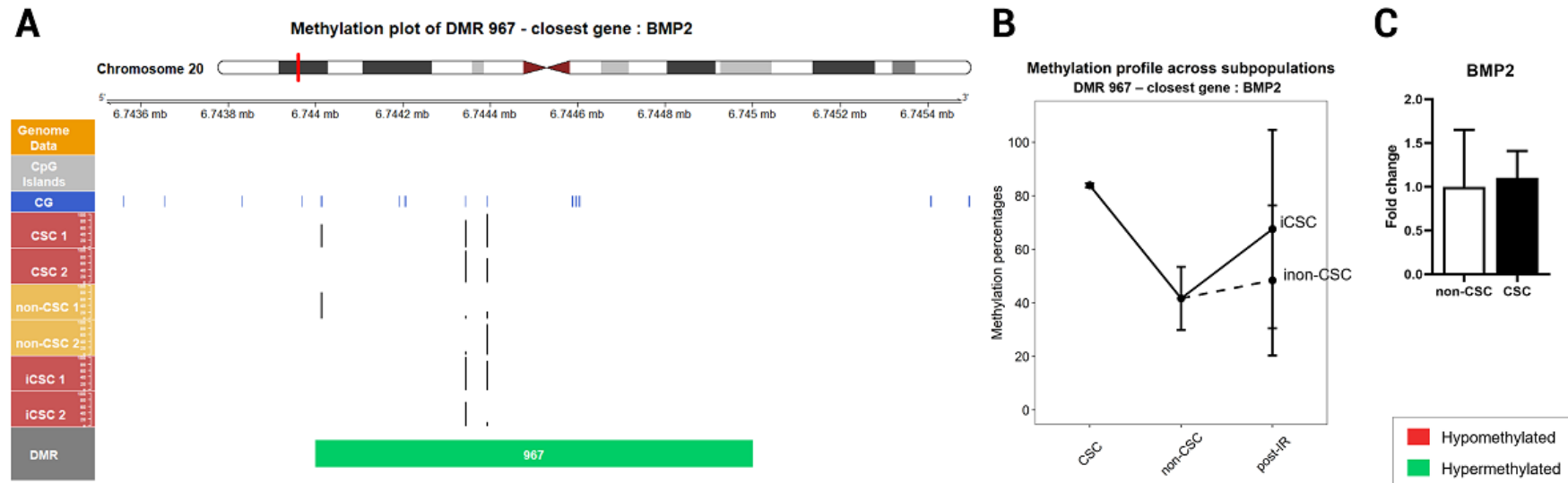


FIGURE 41 METHYLATION PROFILES OF DMRs ASSOCIATED WITH EPHB2, RUNX3, MAPK13, TGFB3, AND BMP2 GENES AND EXPRESSION CORRELATION. **A.** Methylation plots of DMRs close to EPHB2, RUNX3, MAPK13, TGFB3, and BMP2 genes. From top to bottom: chromosome localization; genomic axis; genome data track displaying transcripts along the genome; CpG islands track displaying CGIs along the genome; CG track displaying positions of CpG sites; CSC 1, CSC 2, non-CSC 1, non-CSC 2, iCSC 1 and iCSC 2 tracks displaying CpG methylation percentages histograms; DMR track displaying positions of identified DMRs, hypomethylated ones in CSCs compared to non-CSCs in red, and hypermethylated ones in CSCs compared to non-CSCs in green. **B.** Methylation profiles plots of DMRs close to EPHB2, RUNX3, MAPK13, TGFB3, and BMP2 genes across subpopulations. **C.** Relative mRNA levels of EPHB2, RUNX3, MAPK13, TGFB3 and BMP2 in non-CSC (ALDH⁻) and CSC (ALDH⁺) populations.

Additionally to a lack of functional significance on gene expression for the 5 DMRs tested, this first analysis came with some limitations. Except for the DMR#21 associated with EPHB2, the differential methylation of these 5 DMRs was calculated based on 1 or 2 CpG sites. Differential methylation on 1 or 2 CpG sites does not exclude a functional role in the gene expression regulation, but it reduces the probability to demonstrate both significant methylation differences by an afterward region-specific validation and significant correlation with gene expression. Thus, the absence of a minimum CpG number per region to compute the differential methylation need to be taken into account for the selection of DMRs to validate. Methylation profiles of the 5 selected DMRs did not exactly match the hypothetical pattern of regions that present a CSC-specific methylation level, restored after the conversion from non-CSC to iCSC, but not changing between non-CSC and non-CSC. Therefore, the sorting based on the three pairwise comparisons allows a pre-sorting of regions potentially involved in dedifferentiation but is not sufficient for the selection of the most relevant ones.

In conclusion, this first analysis provided an initial insight into the CpG methylation data obtained by the RRBS analysis, alongside proficiency in sequencing analysis methodology and R programming, as well as parameters and adjustments to overcome the aforementioned limitations. Therefore, this analysis was a first step in the identification of regions regulated by DNA methylation during the dedifferentiation of non-CSCs into iCSCs, which is pursued in a second analysis.

1.2.3 SECOND ANALYSIS OF REDUCED REPRESENTATION BISULFITE SEQUENCING DATA

Using the knowledge provided by the first analysis, I developed a second analysis to go further and find more relevant regions potentially implicated in the radio-induced dedifferentiation, analysis also based on the `methyKit` R package.

The identification of differential methylation at promoter regions and CGIs was not reasonably possible with the use of tile regions as it does not allow the study of specific annotated loci. Hence, the analysis of specific genomic features could also be incorporated into the analysis of these methylation data. As indicated in the workflow from [Figure 40](#), the first main modification in the analysis is the enlargement of the region scope on which the differential methylation is computed. Indeed, additionally to the 1,000 bp tiles subset, new region subsets are incorporated in the analysis: 200 bp tiles, regions surrounding the TSS, CGI regions, and identified methylation segment regions.

To identify new candidates, the filtering based on the methylation profile should be prioritized over the functional relevance of associated genes. Thereby, in this analysis, the identification and selection of regions regulated by DNA methylation changes throughout the dedifferentiation rely more on their methylation profile across subpopulations than in the previous one.

The detailed workflow of the analysis is presented in [Figure 42, panel 1](#). The localization of code listings on the workflow (in red, on the sides) helps situating which step they refer to. The circled

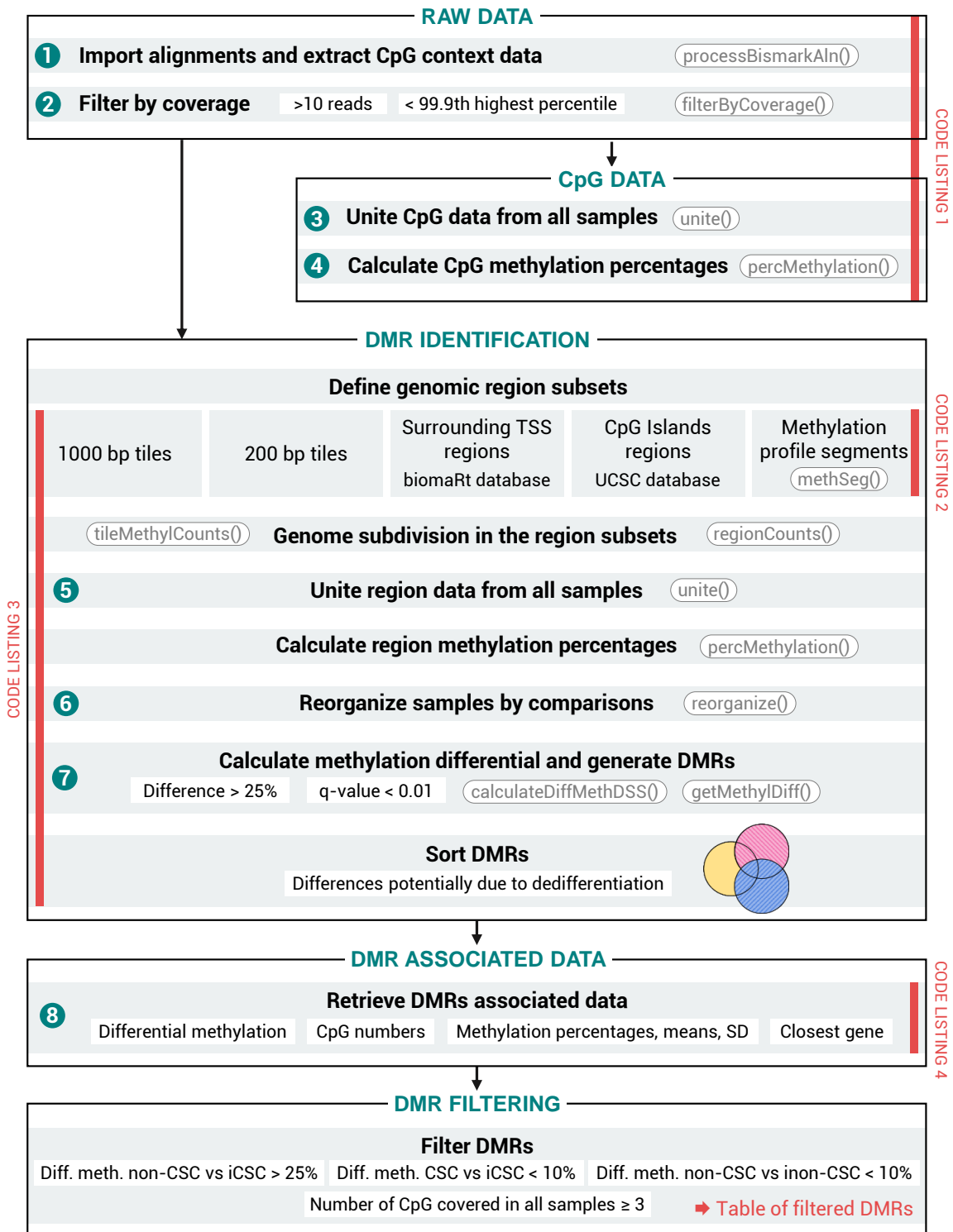


FIGURE 42 Panel 1. Caption in the following pages.

1. DNA METHYLATION CHANGES THROUGHOUT THE RADIO-INDUCED DEDIFFERENTIATION

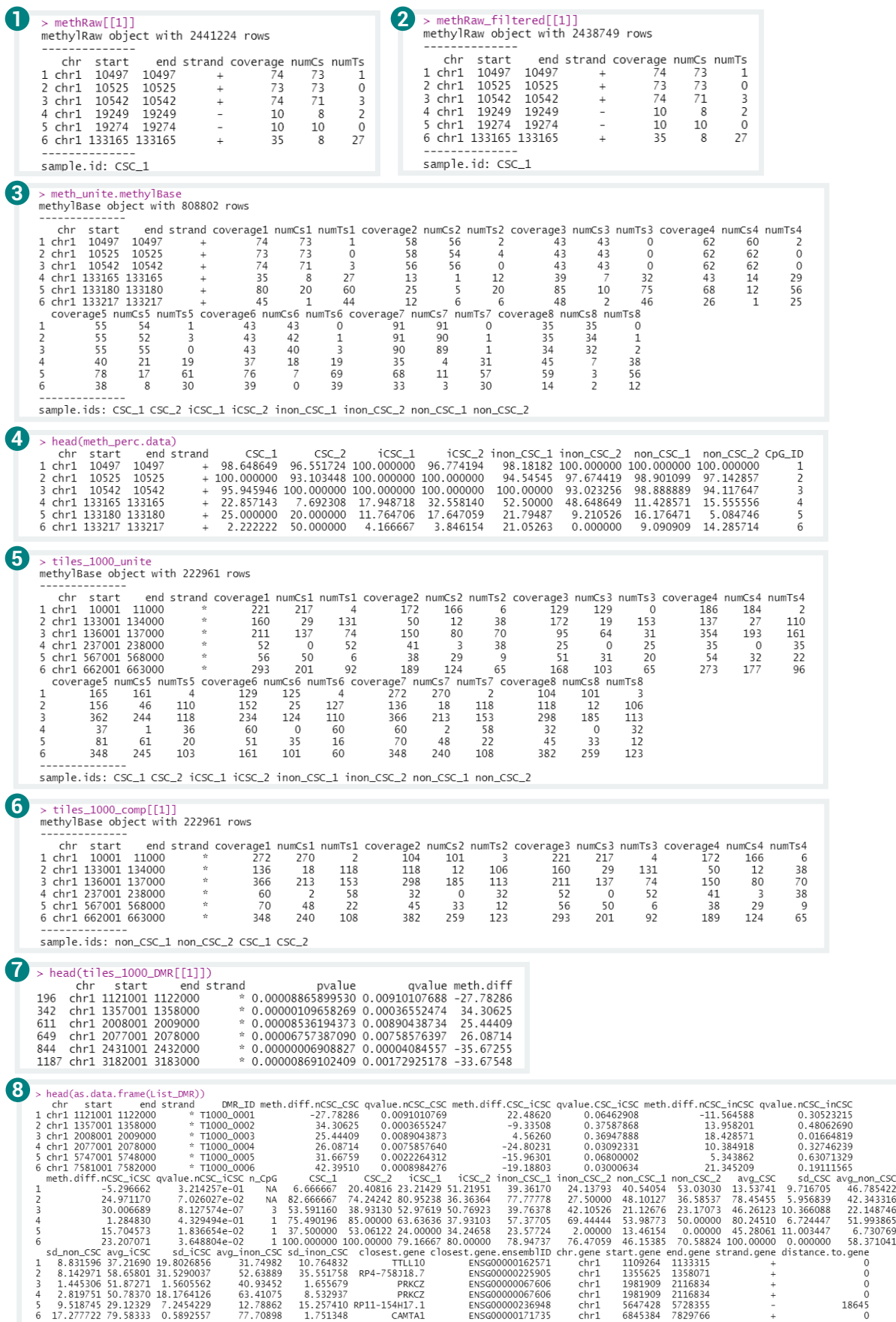


FIGURE 42 Panel 2. Caption next page.

FIGURE 42 RRBS ANALYSIS WORKFLOW AND DATA PROCESSING. **Panel 1.** Workflow of the RRBS data analysis. Circled numbers refer to objects resulting from the corresponding steps displayed in panel 2. Localization of code listings is highlighted in red. Annotated functions come from the `methyKit` R package. **Panel 2.** Associated data objects resulting from the numbered steps of the workflow. ① and ② corresponds respectively to raw and filtered CpG sites read counts (C/T) from the CSC 1 sample only. ③ and ④ corresponds respectively to CpG sites read counts and CpG sites methylation percentages from the 8 samples. ⑤ corresponds to 1,000 bp tiles read counts covered in all samples. ⑥ corresponds to 1,000 bp tiles read counts in non-CSCs and CSCs samples, and ⑦ to differential methylation values (with p- and q- values) computed between those two populations. ⑧ corresponds to the sorted list of 2,596 DMRs with all associated data: methylation differences and q-values, number of CpG sites covered in all samples, methylation percentages in each sample, methylation means per subpopulation with standard deviation and closest genes related information.

numbers from the [Figure 42, panel 1](#) are linked to the ones in [Figure 42, panel 2](#) showing the first rows of objects resulting from each numbered step.

The **raw RRBS data** as read alignment BAM (.bam) files are first imported in R for the 8 samples: CSC 1, CSC 2, iCSC 1, iCSC 2, inon-CSC 1, inon-CSC 2, non-CSC 1 and non-CSC 2, and read counts for cytosines and thymines associated to CpG contexts are retrieved ([Code listing 1](#), [Figure 42, panel 2](#) ②). These read counts per CpG sites coordinates are then **filtered based on their coverage**: CpG sites covered by less than 10 reads in at least one sample are removed, as well as the ones with a coverage superior of the 99.9th percentile, meaning that CpG sites with an aberrantly high coverage value in at least one sample are removed. The first cutoff on minimum read is necessary for statistical robustness, and the second one on high coverage is used to eliminate PCR bias effects ([Code listing 1](#) and [Figure 42, panel 2](#) ②). For example, in the CSC 1 sample, the maximum coverage value for a CpG site is 204,275 reads while the 99.9th of coverage values is 377 reads.

Then, **methylation percentages per CpG sites** are computed based on C and T read counts per CpG. First, data of each sample are merged in one object to restrict the list of CpG to the ones covered in all the 8 samples ([Code listing 1](#) and [Figure 42, panel 2](#) ③ and ④). Additionally, methylation percentages are also computed for all CpG sites of each sample. These CpG methylation percentage data are used to calculate the number of CpG sites per DMR, and for the methylation track display along the genome on methylation plots.

The [Table 15](#) summarizes CpG-associated data obtained in the 8 samples, the number of CpG sites covered per sample, mean of CpG coverage in reads, and mean of methylation percentages, in raw data, filtered data, and united data (CpG sites covered in every sample). The average number of CpG covered before filtering is $2,234,252 \pm 263,552$ CpG sites and after filtering is $2,231,988 \pm 263,288$ CpG sites per sample. 808,802 CpG sites were covered in all 8 samples, corresponding to approximately 36% of the total of filtered CpG sites covered per sample ([Table 15](#)).

To compute the differential methylation on regions and identify DMRs, the genome must be

```

170- ""{r load alignments, eval=F}
171 # Load Bismark RRBS alignment sorted bam
172 aln.files <- as.list(list.files(path = "./alignment", pattern = "*sorted.bam$",
173                               all.files = F,
174                               full.names = T, recursive = F,
175                               ignore.case = F, include.dirs = F))
176
177 # Get methylation percentage from sorted Bismark alignments
178 methRaw <- suppressMessages(
179   processBismarkAln(
180     location = aln.files,
181     sample.id=list("CSC_1", "CSC_2", "iCSC_1", "iCSC_2", "inon_CSC_1", "inon_CSC_2", "non_CSC_1", "non_CSC_2"),
182     read.context="CpG",
183     save.context = c("CpG", "CHG", "CHH"),
184     treatment=c(0,0,1,1,2,2,3,3))
185
186 save(methRaw, file="./data R/methRaw.RData")
187- ""

```

```

196- ""{r ALL : filter by coverage, eval=F}
197 methRaw_filtered <- filterByCoverage(methRaw, lo.count=10, lo.perc=NULL, hi.count=NULL, hi.perc=99.9)
198
199 # Remove controls using the chromosome names
200- for (i in 1:8) {
201   methRaw_filtered[[i]] <- methRaw_filtered[[i]][methRaw_filtered[[i]][["chr"]] %in% chr_levels]
202- }
203
204 save(methRaw_filtered, file="./data R/methRaw_filtered.RData")
205- ""

```

```

222- ""{r CpG : methylation percentages all samples, cache=T, results='hide'}
223 # Unite methylRawList and methylRawListDB objects that only bases with coverage from all samples are retained.
224 meth_unite.methylBase <- suppressMessages(methylKit::unite(methRaw_filtered, destrand=TRUE))
225 # Transform methylBase object in dataframe
226 meth_unite.data <- getData(meth_unite.methylBase)
227
228 # Matrix with percent methylation values per base/region across all samples, row names would be base/region identifiers
229 meth_perc <- as.data.frame(percMethylation(meth_unite.methylBase, rowids = T))
230
231 # Assemble dataframes
232 meth_perc.data <- cbind(meth_unite.data[,1:4], meth_perc)
233
234 # Order by chromosomes levels
235 meth_perc.data <- meth_perc.data[order(factor(meth_perc.data$chr, levels=chr_levels)),]
236
237 # Add CpG ID
238 meth_perc.data <- mutate(meth_perc.data, CpG_ID=row_number())
239- ""

```

```

241- ""{r CpG : methylation percentages individual samples, cache=T, results='hide'}
242 # Get all CpG sites methylation percentages (not only the ones covered in all samples)
243 # Function
244- getPerc <- function(methylRaw.obj){
245   x=getData(methylRaw.obj)
246   x <- mutate(x, perc=100 * numCs/(numCs + numTs))
247   colnames(x)[8] <- methylRaw.obj@sample.id
248   return(x[,c(1,2,3,4,8)])
249- }
250 # Apply function to get CpG methylation percentages of all samples individually
251 meth_perc.all <- list()
252- for (i in 1:length(methRaw_filtered)) {
253   meth_perc.all[[i]] <- getPerc(methRaw_filtered[[i]])
254- }
255- ""

```

CODE LISTING 1 GENERATING CpG METHYLATION PERCENTAGES FROM READ COUNTS.

subdivided into defined regions. Additionally to the previous 1,000 bp tiles subset, new region subsets are incorporated in this new analysis:

- **1,000 bp tiles:** The genome is tiled into 1,000 bp non-overlapping bins, i.e. chr1:1-1,000; chr1:1,001-2,000; chr1:2,001-3,000 (same as in the first analysis).
- **200 bp tiles:** The genome is tiled into 200 bp non-overlapping bins, i.e. chr1:1-200; chr1:201-400; chr1:401-600.
- **Regions surrounding gene transcription start sites (TSSs):** TSSs coordinates on hg19 human genome were retrieved from the BioMart database (`biomaRt` R package), regions from 500 bp upstream to 500 bp downstream (1001 bp long), corresponding to promoter and beginning of gene body, were used.
- **CpG islands regions:** Coordinates of CGI were retrieved from the UCSC database.
- **Methylation segments:** Segments of 10 to 100 adjacent CpG with similar methylation levels.

RESULTS

Sample	Raw CpG data			Filtered CpG data			United CpG data		
	Number of CpG sites	Coverage mean (reads)	Methylation mean (%)	Number of CpG sites	Coverage mean (reads)	Methylation mean (%)	Number of CpG sites	Coverage mean (reads)	Methylation mean (%)
CSC 1	2,441,224	24.83	41.38	2,438,749	23.50	41.38	808,802	40.98	43.71
CSC 2	1,931,696	21.85	41.96	1,929,744	20.64	41.95	808,802	31.38	43.78
iCSC 1	2,170,344	22.11	39.95	2,168,134	20.89	39.94	808,802	33.34	43.10
iCSC 2	2,316,296	23.82	40.54	2,313,947	22.54	40.53	808,802	38.19	43.15
inon-CSC 1	2,258,601	22.40	43.10	2,256,313	21.17	43.10	808,802	34.33	43.07
inon-CSC 2	1,915,478	20.29	42.83	1,913,535	19.13	42.83	808,802	27.99	42.98
non-CSC 1	2,711,453	26.14	42.35	2,708,720	24.77	42.34	808,802	45.84	43.04
non-CSC 2	2,128,926	22.31	41.92	2,126,765	21.06	41.91	808,802	33.36	43.22

TABLE 15 SUMMARY STATISTICS OF CpG NUMBERS, COVERAGE, AND METHYLATION PERCENTAGES IN ALL SAMPLES THROUGHOUT THE ANALYSIS OF RRBS DATA.

To generate **methylation segments**, each sample methylation data is segmented based on methylation profiles using the `methSeg()` function from the `methyKit` R package, relying on a segmentation algorithm provided by the `fastseg` R package ([Code listing 2](#)) (Klambauer *et al.*, 2012). Here the minimum of CpG sites per segment is set to 10. The algorithm incorporates also clustering of segments into 4 groups based on the mean methylation value of each segment. The [Figure 43](#) illustrates the segmentation results for the CSC 1 sample, similar results are obtained for all the samples. Coordinates of identified segments in each sample are then used to aggregate the read counts within their coordinates, to allow the calculation of their differential methylation between subpopulations.

```

951- ""{r SEG : Get segments with methSeg, results='hide', cache=T}
952 # Get segments localizations with CpG with similar methylation, from each samples
953 methSeg_list <- GRangesList()
954 methSeg_summary_plots <- list()
955 methSeg_proportion_plots <- list()
956 for (i in 1:8) {
957   ## Methylation segmentation
958   filename <- paste0("./images/methSeg_summary_",sample_names[i],".png")
959
960   png(filename=filename, width=1000, height=800)
961   methSeg_sample <- methSeg(methRaw_filtered[[i]], minSeg=10, G=1:4, maxInt=100)
962   # minSeg represents the minimal length of segment, meaning the minimal number of CpG in the segment (see fastseg())
963   # maxInt represents the maximal length of the left and the right segment, meaning the maximum number of CpG in the
segment (see fastseg())
964   dev.off()
965
966   methSeg_list[[i]] <- methSeg_sample
967
968   methSeg_summary_plots[[i]] <- filename
969
970   ## Plot of segments methylation proportion
971   filename <- paste0("./images/methSeg_proportion_",sample_names[i],".png")
972
973   png(filename=,
974       width=800, height=400)
975   plot(methSeg_list[[i]]$seg.mean,
976       log10(width(methSeg_list[[i]])),pch=20,
977       col=scales::alpha(methSegplot_colors[as.numeric(methSeg_list[[i]]$seg.group)], 0.2),
978       ylab="log10(length)",
979       xlab="Methylation proportion")
980   dev.off()
981
982   methSeg_proportion_plots[[i]] <- filename
983 }
984 -

```

CODE LISTING 2 GENERATING METHYLATION PROFILE SEGMENTS IN EACH SAMPLE.

For **CpG islands regions**, the UCSC CpG islands database on the hg19 assembly was used, in which a CGI is defined by a GC content > 50%, an observed vs expected ratio higher than 0.6 for the occurrence of CpG sites, and a length greater than 200 bp.

The all process of **DMR identification** is illustrated on the 1,000 bp tiles subset in the [Code list-](#)

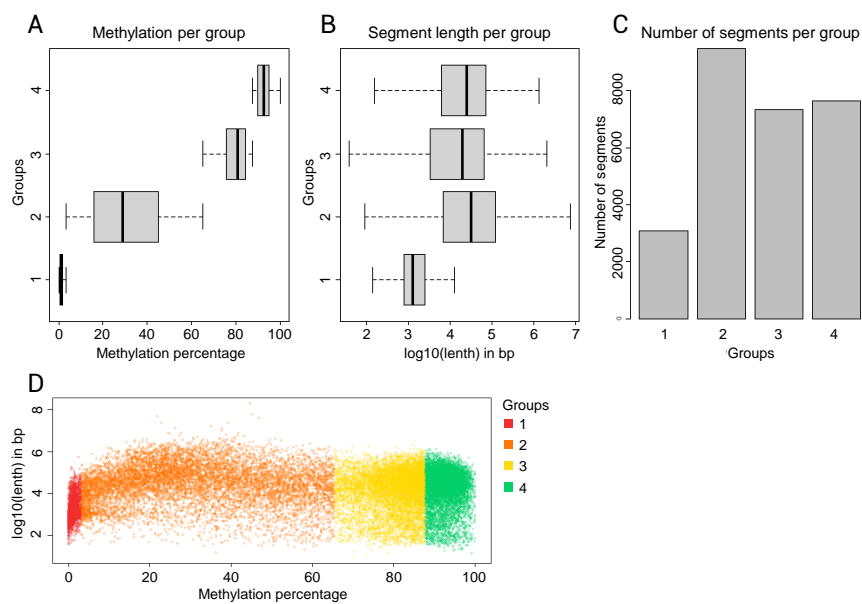


FIGURE 43 METHYLATION SEGMENTATION SUMMARY OF THE CSC 1 SAMPLE. **A.** Methylation percentages distribution per group. **B.** Segment lengths distribution per group. **C.** Number of segments per group. **D.** Distribution of the four segment groups per length and methylation percentages.

ing 3. First, read counts for Cs and Ts are aggregated into corresponding coordinates of subset regions, and regions covered in each sample are kept to calculate each sample's methylation percentages (Code listing 3 and Figure 42, panel 2 (5)). Then, read counts are reorganized in pairwise comparisons of subpopulations and differential methylation of each region is calculated for each comparison (Code listing 3 and Figure 42, panel 2 (6) and (7)).

The **calculation of differential methylation** and its associated significance is processed by the `calculateDiffMethDSS()` function from the `methyKit` R package which come originally from the `DSS` R package (Feng *et al.*, 2022; Feng *et al.*, 2014; Feng and Wu, 2019). The calculation method relies on a beta-binomial model which is the most used model for methylation analysis (Feng *et al.*, 2014). The regions that have a significant differential methylation $> 25\%$ (absolute value, $> 25\%$ for hypermethylated ones, and $< -25\%$ for hypomethylated ones), with a q-value < 0.01 , are considered as differentially methylated between two populations and are referred as DMRs.

In the same way as the first analysis, **DMRs are sorted** to identify the ones potentially involved in the non-CSC radio-induced dedifferentiation into iCSC (Figure 39). DMRs between both CSC states or both non-CSC states are excluded from the CSC-specific DMR pool (significant methylation difference $> 25\%$ between non-CSC and CSC, but $< 25\%$ between both CSC states and both non-CSC states). The sorting of DMRs is carried out on the 5 different regions subsets independently, and results in 5 lists of DMRs: 1,112 DMRs for 1,000 bp tiles, 1,251 DMRs for 200 bp tiles, 109 DMRs surrounding TSS, 21 CGIs and 103 methylation segments, for **a total of 2,596 DMRs identified as having a potential role during the radio-induced phenotypic plasticity from the non-CSC state to the CSC state** (Figure 44).

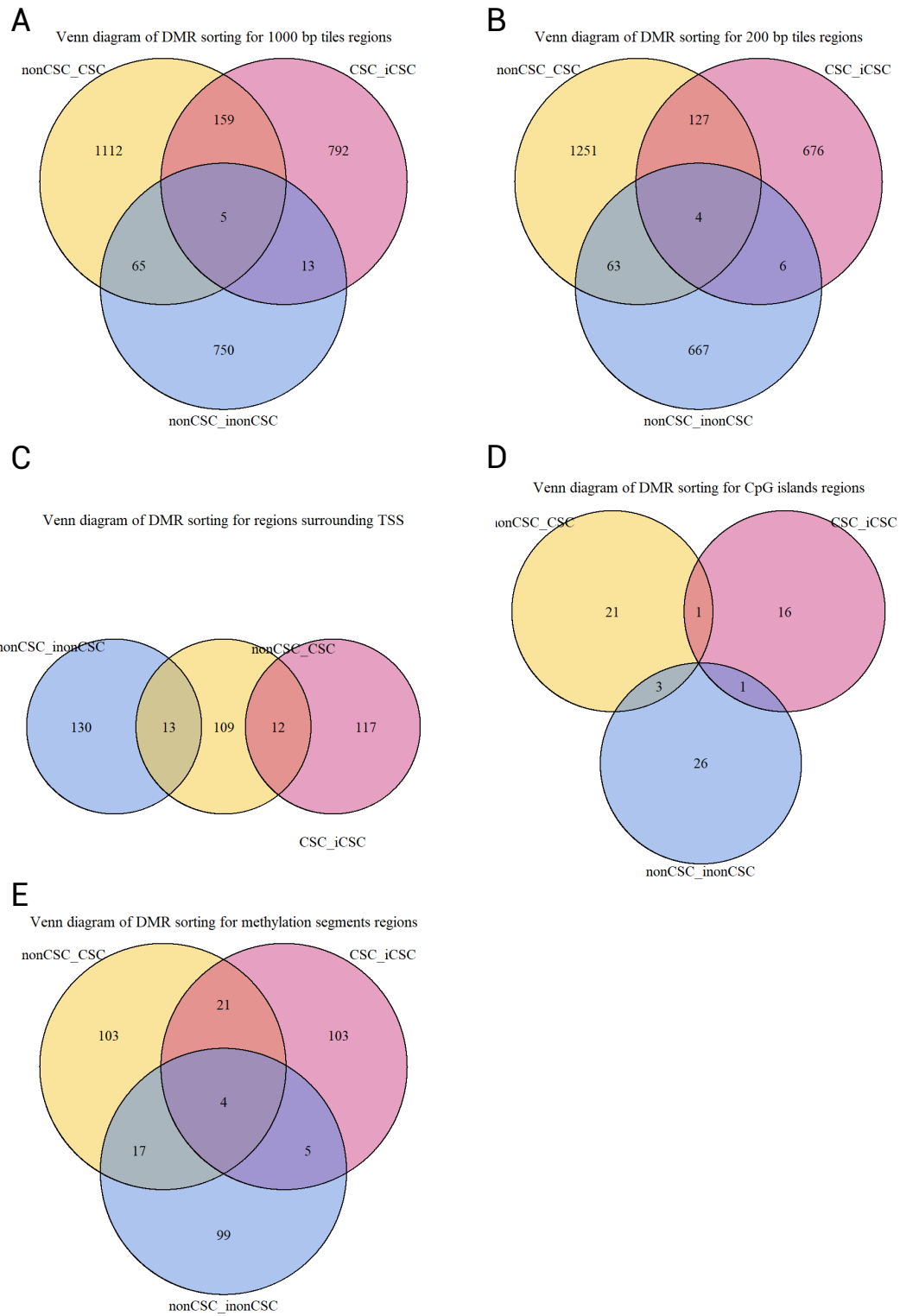


FIGURE 44 VENN DIAGRAMS OF DMR SORTING. Identified DMRs by comparing non-CSC vs CSC, CSC vs iCSC, and non-CSC vs inon-CSC, from the (A) 1,000 bp tiles regions, (B) 200 bp tiles regions, (C) regions surrounding TSS, (D) CpG island regions and (E) methylation segments regions.

```

254 ~~~{r 1000 tiles : UNITE, cache=T, results='hide'}
255 # Summarizes methylated/unmethylated base counts over tiling windows across genome. This function can be used when
# differential methylated analysis is preferable to tiling windows instead of base pairs.
256 tiles_1000 <- tileMethylCounts(methRaw_filtered, win.size=1000, step.size=1000)
257 |
258 tiles_1000_unite <- suppressMessages(methylKit::unite(tiles_1000, destrand=TRUE)) #methylBase object
259
260 tiles_1000_perc <- as.data.frame(percMethylation(tiles_1000_unite, rowids = T)) #percentage methylation in matrix
261
262 write.csv(tiles_1000_perc , file="./data comparative/percMeth_tiles_1000.csv", row.names = T, quote=F)
263 ~~~

268 ~~~{r 1000 tiles : SUBSET, cache=T, results='hide'}
269 # Subset samples in new objects for each comparison
270 tiles_1000_comp <- list()
271 for (i in 1:6) {
272   tiles_1000_comp[[i]] <- reorganize(tiles_1000_unite, sample.ids=sample_comp[[i],treatment=c(0,0,1,1))
273 }
274 ~~~

60 ~~~{r Methylation differential function, include=F}
61
62 MethDiff <- function(x, region){
63
64   # Methylation differential calculation
65   my.diff <- calculateDiffMethDSS(x, adjust="SLIM")
66   my.diff.data <- getData(my.diff)
67
68   # Get DMR : meth diff > 25% & qvalue < 0.01
69   topDiff <- getData(getMethDiff(my.diff, difference = 25, qvalue = 0.01, type = "all"))
70
71   # Save file
72   x_name <- paste(sub('[_]*$', '', x@sample.ids[1]),"vs",sub('[_]*$', '', x@sample.ids[3]), sep="-")
73   write.csv(topDiff,file=paste0("./data comparative/", x_name, "_diff_meth_",region,"_25p_q_0_01.csv"),row.names=F,quote=F)
74
75   return(topDiff)
76 }
77 ~~~

280 ~~~{r 1000 tiles : METH DIFF, cache=T, results='hide'}
281 # Use MethDiff function created to save meth tiles files and return the filtered tiles (difference 25 and q value 0.01)
282 tiles_1000_DMR <- list()
283 for (i in 1:6) {
284   tiles_1000_DMR[[i]] <- MethDiff(tiles_1000_comp[[i]],region="tiles_1000")
285 }
286 ~~~

309 ~~~{r 1000 tiles : SORT, cache=T}
310 # GRanges objects
311 tiles_1000_DMR_GR <- GRangesList()
312 for (i in 1:6) {
313   tiles_1000_DMR_GR[[i]] <- GRanges(tiles_1000_DMR[[i]])
314 }
315
316 # Find overlaps between 1 and 2, 1 and 3
317 overlaps_1000_1_2 <- findOverlaps(tiles_1000_DMR_GR[[1]], tiles_1000_DMR_GR[[2]])
318 overlaps_1000_1_3 <- findOverlaps(tiles_1000_DMR_GR[[1]], tiles_1000_DMR_GR[[3]])
319 # Get all ID of overlapping regions with 1
320 overlaps_1000_all <- c(overlaps_1000_1_2@from, overlaps_1000_1_3@from)
321 rm(overlaps_1000_1_2,overlaps_1000_1_3)
322
323 # Remove overlapping regions
324 tiles_1000_DMR_sorted <- tiles_1000_DMR[[1]][-overlaps_1000_all,]
325
326 # Add DMR ID
327 tiles_1000_DMR_sorted <- mutate(tiles_1000_DMR_sorted, DMR_ID = paste0("T1000_",str_pad(row_number(),4,pad="0")))
328 tiles_1000_DMR_sorted_GR <- GRanges(tiles_1000_DMR_sorted)
329
330 datatable(tiles_1000_DMR_sorted,extensions = 'FixedColumns',options=list(scrollX=TRUE,fixedColumns=TRUE))
331 ~~~

```

CODE LISTING 3 IDENTIFYING DMRs FROM THE 1,000 BP TILES SUBSET.

1.2.4 IDENTIFIED DIFFERENTIALLY METHYLATED REGIONS (DMRs)

Lists of DMRs from the 5 subsets regions are merged to a total of 2,596 DMRs identified as potentially involved in the radio-induced differentiation of breast cancer cells. 1,574 are hypermethylated in CSCs compared to non-CSCs and 1,022 hypomethylated. The coordinates of these regions may overlap, especially for regions surrounding TSS as several TSSs can be annotated for a same gene, and tiles of 200 bp or 1,000 bp can overlap each other or overlap other region subsets. By reducing the list of DMRs to overlapping coordinates, the 2,596 DMRs match 1,916 unique regions in terms of coordinates.

To gather **DMR associated data**, the same process applied for the DMR identification is applied to the 2,596 identified regions ([Code listing 4](#)). To get methylation levels of these regions, read

counts are united for all samples at these region coordinates to compute methylation percentages per sample and methylation mean \pm standard deviation per subpopulation. Differential methylation for the different pairwise comparisons is also generated based on united read counts per region. In addition, based on previously generated CpG methylation data, the number of CpG sites that are covered in all samples per DMR is added and serves as a robustness indicator for the differential methylation calculated. Finally, the closest gene to each DMR is identified, and its coordinates and distance to DMRs information are also added (Code listing 4 and Figure 42, panel 2 ⑧). The 2,596 DMRs are associated with 1,749 unique closest genes.

Among the identified DMRs, **21 CGIs are found differentially methylated**. In particular, a hypomethylated CGI is found within the gene body of the BEST4, BSN, LEMD2, FBXO39, and CDH4 genes, and 1,370 bp upstream of the FSCN1 gene, while a hypermethylated CGI is found within the gene body of the SNRPN, OCA2, NMRK2, GGTL1, and VWFP1 genes.

Moreover, 109 regions surrounding TSS of 81 genes (several TSSs per gene) are found differentially methylated. For instance, the RBP7, RUNX3, VPS26A, CD9, TRAPPC2L, FBXO39, APOC1, MORC2, and TOM1 genes present a hypomethylation around their TSS, while the PCDHGA1, MPZL3, GRAMD1B, MIR7-3HG, and RETN genes have a region surrounding their TSS hypermethylated.

From DMR coordinates, all the **genes within the 2,000 bp range of a DMR** are identified, giving a total of 1,424 genes. By comparing this list with the lists of 209 genes annotated in the “stem cell population maintenance” gene ontology (GO:0019827), 11 genes are identified: BMP7, FANCC, MED12, NR2E1, PRDM16, SFRP1, STAG2, STAT3, WNT7A, WNT9B, and ZHX2. Among them, FANCC, NR2E1, SFRP1, STAG2, STAT3, and WNT7A are associated with at least one hypomethylated DMR and BMP7, MED12, PRDM16, STAG2, WNT9B, and ZHX2 are associated with at least one hypermethylated DMR (in CSCs compared to non-CSCs).

From the list of 1,424 genes in the 2 kb range of DMRs, the ones with at least one hypomethylated and one hypermethylated DMR within the 2 kb range are excluded, leaving a total of 1,374 genes with a consistent differential methylation in the case of several proximal DMRs. This list is separated into two subsets, the ones associated with at least one hypomethylated DMR (602 genes) and the ones associated with at least one hypermethylated DMR (780 genes) to carry out the **pathway enrichment analysis** (Table 16). We can notice the Hedgehog pathway in the top enriched pathways, containing the DHH, ARNTL, and GLI2 genes associated, each one with a hypomethylated DMR in CSCs.

As seen previously with the first analysis, the methylation profile of DMRs across subpopulations is a key feature to find regions undergoing coherent methylation changes throughout dedifferentiation. So, to find the most susceptible ones to participate in the dedifferentiation process, **DMRs are selected based on their methylation profiles**. As illustrated in Figure 45, in addition to the

```

1152 ~~~{r GENERATE METHYLATION DATA ON DMR POSITIONS}
1153 # Assemble all DMR subsets :
1154 List_DMR <- rbind(tiles_1000_DMR_sorted,
1155                 tiles_200_DMR_sorted,
1156                 surroundTSS_DMR_sorted,
1157                 CGIregions_DMR_sorted,
1158                 SEGregions_DMR_sorted)
1159 # Manage columns
1160 List_DMR <- List_DMR[,c("chr", "start", "end", "strand", "DMR_ID", "meth.diff", "qvalue")]
1161 colnames(List_DMR)[6:7] <- c("meth.diff.nCSC_CSC", "qvalue.nCSC_CSC")
1162
1163 # Create Genomic Ranges object from DMRs localizations
1164 List_DMR_GR <- GRanges(List_DMR)
1165
1166 # Subset filtered methylation data into DMR localizations
1167 DMR_regions <- regionCounts(methRaw_filtered, List_DMR_GR)
1168
1169 # Unite methylation data covered in all samples
1170 DMR_regions_unite <- suppressMessages(methylKit::unite(DMR_regions, destrand=TRUE))
1171 DMR_regions_unite.data <- getData(DMR_regions_unite)
1172 ~~~

1174 ~~~{r GET DIFFERENTIAL METHYLATION}
1175 # Reorganize by sample comparisons
1176 ## sample comparison 2 : CSC vs iCSC
1177 DMR_comp_2 <- reorganize(DMR_regions_unite, sample.ids=sample_comp[[2]], treatment=c(0,0,1,1))
1178 ## sample comparison 3 : non-CSC vs inon-CSC
1179 DMR_comp_3 <- reorganize(DMR_regions_unite, sample.ids=sample_comp[[3]], treatment=c(0,0,1,1))
1180 ## sample comparison 4 : non-CSC vs iCSC
1181 DMR_comp_4 <- reorganize(DMR_regions_unite, sample.ids=sample_comp[[4]], treatment=c(0,0,1,1))
1182
1183 # Calculate differential methylation for all comparisons
1184 DMR_comp_2 <- getData(calculatedDiffMethDSS(DMR_comp_2, adjust="SLIM"))
1185 DMR_comp_3 <- getData(calculatedDiffMethDSS(DMR_comp_3, adjust="SLIM"))
1186 DMR_comp_4 <- getData(calculatedDiffMethDSS(DMR_comp_4, adjust="SLIM"))
1187
1188 # Rename columns
1189 colnames(DMR_comp_2)[c(6,7)] <- c("qvalue.CSC_iCSC", "meth.diff.CSC_iCSC")
1190 colnames(DMR_comp_3)[c(6,7)] <- c("qvalue.nCSC_inCSC", "meth.diff.nCSC_inCSC")
1191 colnames(DMR_comp_4)[c(6,7)] <- c("qvalue.nCSC_iCSC", "meth.diff.nCSC_iCSC")
1192 ~~~

1194 ~~~{r GET METHYLATION PERCENTAGES}
1195 # Compute methylation percentages on regions corresponding to DMRs
1196 DMR_regions_perc <- as.data.frame(percMethylation(DMR_regions_unite, rowids = T))
1197 DMR_regions_perc <- cbind(DMR_regions_unite.data[,c(1:4)], DMR_regions_perc)
1198
1199 # Compute methylation percentages means and SD
1200 DMR_regions_perc <- DMR_regions_perc %>%
1201   mutate(avg_CSC = rowMeans(DMR_regions_perc[,c("CSC_1", "CSC_2")], na.rm=T),
1202          sd_CSC = rowSds(as.matrix(DMR_regions_perc[,c("CSC_1", "CSC_2")]), na.rm=T),
1203          avg_non_CSC = rowMeans(DMR_regions_perc[,c("non_CSC_1", "non_CSC_2")], na.rm=T),
1204          sd_non_CSC = rowSds(as.matrix(DMR_regions_perc[,c("non_CSC_1", "non_CSC_2")]), na.rm=T),
1205          avg_iCSC = rowMeans(DMR_regions_perc[,c("iCSC_1", "iCSC_2")], na.rm=T),
1206          sd_iCSC = rowSds(as.matrix(DMR_regions_perc[,c("iCSC_1", "iCSC_2")]), na.rm=T),
1207          avg_inon_CSC = rowMeans(DMR_regions_perc[,c("inon_CSC_1", "inon_CSC_2")], na.rm=T),
1208          sd_inon_CSC = rowSds(as.matrix(DMR_regions_perc[,c("inon_CSC_1", "inon_CSC_2")]), na.rm=T))
1209 ~~~

1211 ~~~{r ASSOCIATE CpG DATA TO DMR}
1212 # CpG data as GRanges object
1213 meth_perc_GR <- GRanges(meth_perc.data)
1214 # Find overlaps between sorted DMR and CpGs
1215 overlaps_CpG <- findOverlaps(meth_perc_GR, List_DMR_GR)
1216 # Keep only CpG overlapping with a DMR
1217 CpG_in_DMR <- meth_perc.data[overlaps_CpG@from,]
1218 # Add DMR_ID column to CpG data
1219 CpG_in_DMR <- mutate(CpG_in_DMR, DMR_ID = deframe(List_DMR[overlaps_CpG@to, "DMR_ID"]))
1220
1221 # Get number of CpG
1222 nCpG_in_DMR <- CpG_in_DMR %>%
1223   group_by(DMR_ID) %>%
1224   summarise(n_CpG = n())
1225 ~~~

1227 ~~~{r JOIN ALL ASSOCIATED DMR DATA}
1228 # Differential methylation
1229 List_DMR <- left_join(List_DMR, DMR_comp_2[,c(1,2,3,4,7,6)], by=c("chr", "start", "end", "strand"))
1230 List_DMR <- left_join(List_DMR, DMR_comp_3[,c(1,2,3,4,7,6)], by=c("chr", "start", "end", "strand"))
1231 List_DMR <- left_join(List_DMR, DMR_comp_4[,c(1,2,3,4,7,6)], by=c("chr", "start", "end", "strand"))
1232
1233 # Number of CpG
1234 List_DMR <- left_join(List_DMR, nCpG_in_DMR, by=c("DMR_ID"))
1235
1236 # Methylation percentages
1237 List_DMR <- left_join(List_DMR, DMR_regions_perc, by=c("chr", "start", "end", "strand"))
1238
1239 # Closest gene
1240 List_DMR <- List_DMR %>%
1241   mutate(closest.gene = genes_data[nearest(List_DMR_GR, genes_data_GR), "external_gene_name"],
1242          closest.gene.ensemblID = genes_data[nearest(List_DMR_GR, genes_data_GR), "ensembl_gene_id"],
1243          chr.gene = genes_data[nearest(List_DMR_GR, genes_data_GR), "chr"],
1244          start.gene = genes_data[nearest(List_DMR_GR, genes_data_GR), "start"],
1245          end.gene = genes_data[nearest(List_DMR_GR, genes_data_GR), "end"],
1246          strand.gene = genes_data[nearest(List_DMR_GR, genes_data_GR), "strand"],
1247          distance.to.gene = distanceToNearest(List_DMR_GR, genes_data_GR)$elementMetadata$distance)
1248 ~~~

```

CODE LISTING 4 ASSOCIATING DMRs WITH THEIR RELATED DATA.

RESULTS

#	Term	Overlap	P-value	Odds Ratio	Combined Score	Genes associated to hypomethylated DMRs in CSCs
1	FTO Obesity Variant Mechanism	2/8	0.022	10.77	40.90	FTO; PRDM16
2	Role of <i>Osx</i> and miRNAs in tooth development	3/15	0.009	8.09	37.74	NOTCH3; ALPL; BMP7
3	Hedgehog Signaling Pathway Netpath	3/16	0.011	7.47	33.45	DHH; ARNTL; GLI2
4	miRNA targets in ECM and membrane receptors	3/22	0.027	5.11	18.39	COL5A1; COL4A1; COL5A2
5	PKC-gamma calcium signaling pathway in ataxia	3/22	0.027	5.11	18.39	PLCB4; TRPC3; ATP2B2
6	MET in type 1 papillary renal cell carcinoma	6/59	0.008	3.67	17.53	ALK; MAP2K2; PTPN11; ETS1; CRKL; PAK4
7	Development of ureteric collection system	5/47	0.013	3.86	16.72	BMP4; FRAS1; GFRA1; CELSR1; GLI2
8	Somatic sex determination	2/14	0.065	5.39	14.73	NR5A1; DHH
9	Differentiation of white and brown adipocyte	3/25	0.038	4.41	14.40	BMP4; PRDM16; BMP7
10	Nuclear receptors	4/38	0.027	3.81	13.80	NR5A1; RXRA; NR1H2; PGR

#	Term	Overlap	P-value	Odds Ratio	Combined Score	Genes associated to hypermethylated DMRs in CSCs
1	The alternative pathway of fetal androgen synthesis	3/11	0.008	9.27	45.11	POR; HSD17B3; HSD17B6
2	Globo Sphingolipid Metabolism	4/21	0.008	5.82	28.05	ST6GALNAC1; ST6GAL1; GCNT1; ABO
3	Development and heterogeneity of the ILC family	5/32	0.007	4.59	22.45	IL33; BCL11B; TBX21; ID2; RORA
4	Cell Differentiation - Index expanded	3/19	0.036	4.63	15.41	MYOD1; ID2; STAT3
5	Vitamin D-sensitive calcium signaling in depression	5/41	0.021	3.44	13.30	GGTLC1; KCNQ2; ITPR1; ATP2B3; SLC8A1
6	Splicing factor NOVA regulated synaptic proteins	5/42	0.023	3.34	12.62	NTNG1; CHL1; KCNQ2; PRKCG; NEO1
7	Cell Differentiation - Index	2/13	0.089	4.49	10.85	MYOD1; STAT3
8	Development of pulmonary dendritic cells and macrophage subse	2/13	0.089	4.49	10.85	ID2; STAT3
9	Pathways Regulating Hippo Signaling	9/98	0.014	2.51	10.66	PDGFRB; NTRK2; PRKCH; GNAL; FLT1; FLT3; GNAS; PRKCG; CDH7
10	TCA Cycle Nutrient Utilization and Invasiveness of Ovarian Cancr	1/5	0.180	6.17	10.56	STAT3

TABLE 16 PATHWAYS ENRICHMENT IN GENES ASSOCIATED WITH IDENTIFIED DMRs. The top 10 enriched pathways (sorted by combined score, computed based on p-value and odds ratio) in genes within the 2 kb range of identified DMRs, either hypomethylated ones (top table) or hypermethylated ones (bottom table) (WikiPathways Human 2021 database).

previously used 3 pairwise comparisons, a fourth one is also taken into account: non-CSC vs iCSC, corresponding directly to methylation changes occurring during dedifferentiation. Hence, DMRs that present a methylation profile close to the hypothetical, one represented in [Figure 45](#) (for hypomethylated DMRs, reverse profile for hypermethylated ones), are selected thanks to these filters: differential methylation between non-CSC and CSC > 25%, differential methylation between non-CSC and iCSC > 25%, a differential between CSC and iCSC < 10% and a differential between non-CSC and inon-CSC < 10%. Compared to the previously identified DMRs that were sorted only based on differential methylation greater or lower than 25% (non-CSC vs CSC > 25% while differences between the two CSC or two non-CSC states < 25%), selected DMRs with these additional parameters have a methylation profile that fits the hypothetical one, meaning they have greater probabilities to be involved in phenotypic switches from non-CSCs to iCSCs.

In addition, to add robustness to differential methylation values, only the DMRs containing at least 3 CpG sites covered in all samples are kept to select DMRs for validation ([Figure 42, panel 1](#)). Indeed, as previously seen in the first analysis, filtering DMRs based on a minimal number of CpG sites should help to find significant methylation differences in afterward region-specific validation, as well as significant correlation with associated gene expression.

From the 2,596 identified DMRs, **35 DMRs are selected** based on their dedifferentiation-associated methylation profile and with a sufficient number of CpG sites covered in all samples. The list of these 35 selected DMRs, associated with 23 unique genes, along with their methylation levels and differences are displayed in [Table 17](#). Among them, one CGI differentially methylated is selected, located at 1,370 bp upstream of the *FSCN1* gene, and one DMR surrounding a TSS is

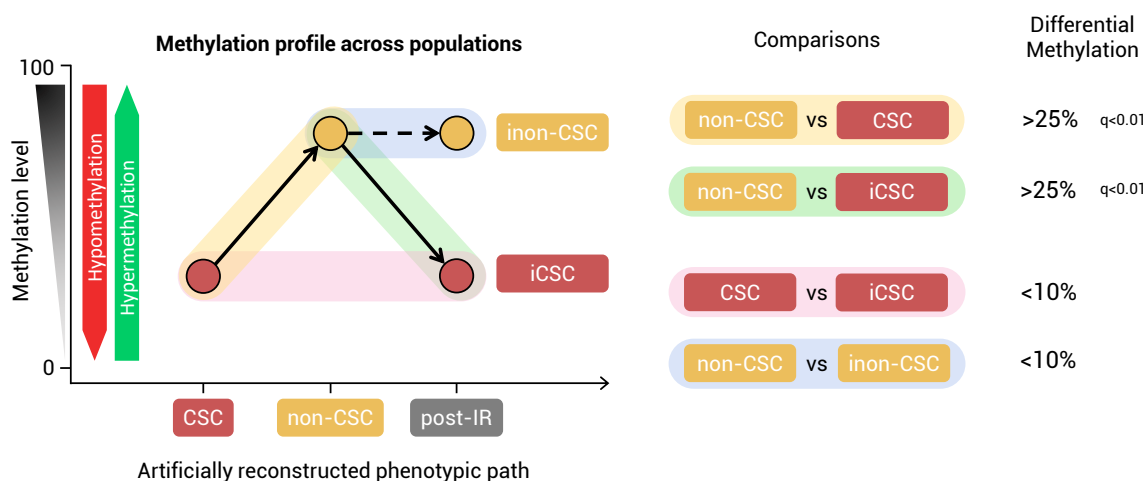


FIGURE 45 FILTERING DMRs BASED ON THEIR METHYLATION PROFILES INDICATING POTENTIAL INVOLVEMENT IN DEDIFFERENTIATION. Hypothetical methylation profile of a region that could be involved in radio-induced dedifferentiation through the DNA methylation regulation of its associated gene. This DMR is hypomethylated in CSCs compared to non-CSCs as well as in iCSCs compared to non-CSCs, and no differences are found between the two CSC states and the two non-CSC states. The reverse pattern can be applied to DMRs hypermethylated in CSCs compared to non-CSCs. To select DMRs matching this methylation profile pattern, they are filtered based on the differential methylation between subpopulations (absolute values), differences need to be greater than 25%, and similarities are defined by less than 10% methylation difference.

selected, associated with the CD9 gene. CpG methylation levels and methylation profiles of 5 of these DMRs, associated with fascin actin-bundling protein 1 (FSCN1), cholinergic receptor nicotinic alpha 6 subunit (CHRNA6), cadherin 7 (CDH7), CD9, and protein kinase CAMP-dependent type I regulatory subunit beta (PRKAR1B) genes are available in [Figure 46](#).

In conclusion, a list of 35 DMRs, including 5 particularly interesting ones, have been identified and need to be further validated. The use of a region-specific DNA methylation quantification technique is required to confirm methylation differences across the four studied subpopulations. Additionally, the correlation of these methylation differences with the expression changes of associated genes should also be assessed.

As the logical continuation of this methylation study was to validate methylation differences of DMRs, the bisulfite sequencing PCR (BSP) approach with the direct sequencing of PCR products (direct-BSP) was chosen to quantify region-specific methylation levels. After a first test, the lack of a practical tool to analyze direct-BSP data led to the development of a new analysis tool, described in the next chapter.

#	DMR ID	Chr.	Start	End	Length	CpG	CSC methylation Mean ± SD	non-CSC methylation Mean ± SD	iCSC methylation Mean ± SD	inon-CSC methylation Mean ± SD	Closest gene	Distance to gene
1	T1000_0078	chr1	237235001	237236000	1000	3	11.62 ± 5.40	42.41 ± 10.73	11.56 ± 11.38	39.92 ± 5.22	RYR2	0
2	T1000_0179	chr3	71478001	71479000	1000	4	38.95 ± 0.05	67.92 ± 2.22	42.00 ± 0.09	72.72 ± 7.04	FOXP1	0
3	T1000_0260	chr5	3001001	3002000	1000	3	63.41 ± 2.64	31.24 ± 2.19	57.94 ± 3.22	33.83 ± 0.25	RP11-3507.1	33258
4	T1000_0293	chr5	149978001	149979000	1000	4	31.94 ± 5.93	56.73 ± 6.12	19.70 ± 7.39	57.23 ± 6.01	SYNPO	1641
5	T1000_0322	chr6	29831001	29832000	1000	3	55.59 ± 0.72	28.34 ± 6.41	55.71 ± 2.90	39.28 ± 7.61	MICF	10732
6	T1000_0360	chr6	162294001	162295000	1000	3	23.64 ± 5.14	67.80 ± 4.04	29.29 ± 3.07	61.57 ± 6.73	PARK2	0
7	T1000_0368	chr6	169249001	169250000	1000	3	19.69 ± 4.00	55.59 ± 4.19	22.37 ± 0.59	60.71 ± 13.13	RP3-495K2.3	112704
8	T1000_0372	chr7	579001	580000	1000	8	23.03 ± 10.72	57.16 ± 5.73	25.87 ± 2.89	53.08 ± 1.24	PRKAR1B	8833
9	T1000_0373	chr7	996001	997000	1000	4	28.82 ± 7.18	56.16 ± 3.04	30.58 ± 6.09	55.11 ± 2.88	COX19	0
10	T1000_0638	chr11	118133001	118134000	1000	3	82.41 ± 3.12	49.62 ± 11.07	80.71 ± 2.95	59.19 ± 27.14	MPZL2	0
11	T1000_0777	chr15	77889001	77890000	1000	3	78.82 ± 8.61	45.17 ± 4.21	82.95 ± 11.25	42.48 ± 18.49	RP11-307C19.2	0
12	T1000_0830	chr16	81481001	81482000	1000	3	27.15 ± 7.15	55.54 ± 4.04	24.59 ± 4.64	49.05 ± 13.12	CMIP	0
13	T1000_0912	chr18	76725001	76726000	1000	7	49.11 ± 4.30	22.95 ± 0.35	50.22 ± 0.49	25.37 ± 1.59	RP11-849I19.1	10554
14	T1000_0970	chr20	19908001	19909000	1000	3	48.93 ± 6.35	83.26 ± 10.06	41.84 ± 14.17	88.78 ± 11.83	RIN2	0
15	T1000_0986	chr20	38764001	38765000	1000	7	48.03 ± 8.43	21.38 ± 3.63	47.68 ± 2.96	28.06 ± 13.51	RP1-191L6.2	9002
16	T1000_1078	chrX	56830001	56831000	1000	3	29.80 ± 0.49	61.74 ± 8.43	26.09 ± 7.15	65.60 ± 15.52	RP11-622K12.1	0
17	T200_0322	chr5	3001201	3001400	200	3	63.41 ± 2.64	31.24 ± 2.19	57.94 ± 3.22	33.83 ± 0.25	RP11-3507.1	33458
18	T200_0431	chr6	169249001	169249200	200	3	19.69 ± 4.00	55.59 ± 4.19	22.37 ± 0.59	60.71 ± 13.13	RP3-495K2.3	113504
19	T200_0439	chr7	579201	579400	200	7	29.72 ± 8.97	66.15 ± 5.99	35.35 ± 8.31	64.83 ± 3.76	PRKAR1B	9433
20	T200_0448	chr7	5647801	5648000	200	3	33.71 ± 15.53	69.11 ± 3.07	40.54 ± 3.98	72.92 ± 15.99	FSCN1	1514
21	T200_0612	chr9	136529401	136529600	200	3	32.92 ± 8.08	55.68 ± 0.48	30.68 ± 0.39	44.10 ± 14.68	SARDH	0
22	T200_0794	chr12	133613401	133613600	200	3	32.19 ± 13.03	74.47 ± 9.69	32.54 ± 4.23	69.11 ± 10.98	RP11-386I8.6	0
23	T200_0871	chr15	77889801	77890000	200	3	78.82 ± 8.61	45.17 ± 4.21	82.95 ± 11.25	42.48 ± 18.49	RP11-307C19.2	0
24	T200_0937	chr16	81481001	81481200	200	3	27.15 ± 7.15	55.54 ± 4.04	25.39 ± 3.50	49.05 ± 13.12	CMIP	0
25	T200_1030	chr19	1094001	1094200	200	3	46.69 ± 6.49	78.97 ± 7.24	49.62 ± 9.16	73.03 ± 2.51	POLR2E	0
26	T200_1089	chr20	19908801	19909000	200	3	48.93 ± 6.35	81.60 ± 7.71	41.84 ± 14.17	84.87 ± 16.68	RIN2	0
27	T200_1099	chr20	38764401	38764600	200	7	48.03 ± 8.43	21.38 ± 3.63	47.68 ± 2.96	28.06 ± 13.51	RP1-191L6.2	9402
28	T200_1205	chrX	56830001	56830200	200	3	29.80 ± 0.49	61.74 ± 8.43	26.09 ± 7.15	65.60 ± 15.52	RP11-622K12.1	0
29	STSS_0055	chr12	6308381	6309381	1001	5	6.79 ± 3.25	46.68 ± 6.75	11.96 ± 7.60	41.42 ± 34.77	CD9	0
30	CGIR_0005	chr7	5647657	5647984	328	8	27.42 ± 5.09	61.42 ± 5.63	34.23 ± 2.28	65.66 ± 10.70	FSCN1	1370
31	SEGR_0019	chr3	71478053	71478253	201	4	38.45 ± 0.75	67.49 ± 0.86	37.10 ± 0.57	71.68 ± 5.56	FOXP1	0
32	SEGR_0020	chr3	71478053	71478274	222	4	38.95 ± 0.05	67.92 ± 2.22	42.00 ± 0.09	72.72 ± 7.04	FOXP1	0
33	SEGR_0021	chr3	71478171	71478274	104	3	38.95 ± 0.05	71.32 ± 0.44	43.56 ± 2.29	76.58 ± 1.59	FOXP1	0
34	SEGR_0040	chr8	42623718	42641148	17431	5	48.23 ± 3.28	17.00 ± 2.13	45.94 ± 3.73	18.48 ± 5.35	CHRNA6	0
35	SEGR_0087	chr18	63418129	63418179	51	5	89.22 ± 4.96	55.95 ± 7.55	80.41 ± 4.40	63.21 ± 5.95	CDH7	0

TABLE 17 Panel 1. Caption in following pages.

#	DMR ID	Chr.	Start	End	Length	CpG	non-CSC vs CSC		non-CSC vs iCSC		CSC vs iCSC		non-CSC vs non-CSC		Closest gene	Distance to gene
							Diff. Meth.	q-value	Diff. Meth.	q-value	Diff. Meth.	q-value	Diff. Meth.	q-value		
1	T1000_0078	chr1	237235001	237236000	1000	3	-26.41	3.63E-03	-25.54	4.79E-03	0.87	5.70E-01	1.48	7.20E-01	RYR2	0
2	T1000_0179	chr3	71478001	71479000	1000	4	-28.49	1.20E-04	-25.41	1.28E-05	3.09	4.18E-01	4.93	4.94E-01	FOXP1	0
3	T1000_0260	chr5	3001001	3002000	1000	3	32.04	4.55E-05	26.80	5.01E-05	-5.24	3.26E-01	2.98	6.09E-01	RP11-3507.1	33258
4	T1000_0293	chr5	149978001	149979000	1000	4	-25.25	3.24E-04	-34.22	3.29E-07	-8.97	1.48E-01	1.24	7.20E-01	SYNPO	1641
5	T1000_0322	chr6	29831001	29832000	1000	3	25.58	4.30E-03	25.39	2.23E-04	-0.19	6.02E-01	8.73	3.51E-01	MICF	10732
6	T1000_0360	chr6	162294001	162295000	1000	3	-43.61	9.52E-07	-37.96	1.82E-08	5.65	3.55E-01	-4.46	5.35E-01	PARK2	0
7	T1000_0368	chr6	169249001	169250000	1000	3	-35.51	2.28E-07	-33.04	2.47E-06	2.47	4.64E-01	4.08	6.40E-01	RP3-495K2.3	112704
8	T1000_0372	chr7	579001	580000	1000	8	-31.23	6.15E-04	-30.40	7.29E-09	0.82	5.67E-01	-3.54	5.31E-01	PRKAR1B	8833
9	T1000_0373	chr7	996001	997000	1000	4	-26.19	5.26E-04	-25.85	3.54E-05	0.34	5.94E-01	-1.47	6.99E-01	COX19	0
10	T1000_0638	chr11	118133001	118134000	1000	3	33.53	6.22E-04	31.96	5.81E-05	-1.57	5.06E-01	8.97	5.59E-01	MPZL2	0
11	T1000_0777	chr15	77889001	77890000	1000	3	35.86	1.61E-03	40.76	2.05E-05	4.90	3.93E-01	-2.89	7.06E-01	RP11-307C19.2	0
12	T1000_0830	chr16	81481001	81482000	1000	3	-27.89	1.68E-03	-32.29	2.80E-06	-4.41	3.63E-01	-7.80	4.63E-01	CMIP	0
13	T1000_0912	chr18	76725001	76726000	1000	7	26.22	2.43E-05	27.01	4.82E-07	0.79	5.58E-01	2.35	6.08E-01	RP11-849I19.1	10554
14	T1000_0970	chr20	19908001	19909000	1000	3	-35.41	2.06E-04	-42.11	1.26E-05	-6.70	3.53E-01	0.73	7.55E-01	RIN2	0
15	T1000_0986	chr20	38764001	38765000	1000	7	26.51	5.88E-04	25.22	1.23E-06	-1.29	5.36E-01	5.84	5.10E-01	RP1-191L6.2	9002
16	T1000_1078	chrX	56830001	56831000	1000	3	-29.74	1.21E-04	-34.53	1.01E-06	-4.79	3.08E-01	3.83	6.55E-01	RP11-622K12.1	0
17	T200_0322	chr5	3001201	3001400	200	3	32.04	6.69E-05	26.80	5.01E-05	-5.24	3.26E-01	2.98	6.09E-01	RP11-3507.1	33458
18	T200_0431	chr6	169249001	169249200	200	3	-35.51	5.22E-07	-33.04	2.47E-06	2.47	4.64E-01	4.08	6.40E-01	RP3-495K2.3	113504
19	T200_0439	chr7	579201	579400	200	7	-33.44	5.32E-05	-29.31	1.64E-05	4.13	3.99E-01	-1.03	7.20E-01	PRKAR1B	9433
20	T200_0448	chr7	5647801	5648000	200	3	-33.85	4.55E-03	-28.12	2.32E-06	5.73	3.65E-01	3.64	6.46E-01	FSCN1	1514
21	T200_0612	chr9	136529401	136529600	200	3	-25.77	6.31E-03	-25.04	5.53E-05	0.74	5.71E-01	-9.34	4.14E-01	SARDH	0
22	T200_0794	chr12	133613401	133613600	200	3	-35.48	5.77E-03	-39.32	5.38E-08	-3.84	4.33E-01	-2.71	6.77E-01	RP11-386I8.6	0
23	T200_0871	chr15	77889801	77890000	200	3	35.86	2.82E-03	40.76	2.05E-05	4.90	3.93E-01	-2.89	7.06E-01	RP11-307C19.2	0
24	T200_0937	chr16	81481001	81481200	200	3	-27.89	2.78E-03	-30.94	1.03E-05	-3.05	4.40E-01	-7.80	4.63E-01	CMIP	0
25	T200_1030	chr19	1094001	1094200	200	3	-30.30	6.46E-03	-28.80	1.60E-03	1.50	5.52E-01	-4.41	5.57E-01	POLR2E	0
26	T200_1089	chr20	19908801	19909000	200	3	-32.52	7.87E-04	-39.22	2.56E-05	-6.70	3.53E-01	-0.61	7.62E-01	RIN2	0
27	T200_1099	chr20	38764401	38764600	200	7	26.51	1.15E-03	25.22	1.23E-06	-1.29	5.36E-01	5.84	5.10E-01	RP1-191L6.2	9402
28	T200_1205	chrX	56830001	56830200	200	3	-29.74	2.39E-04	-34.53	1.01E-06	-4.79	3.08E-01	3.83	6.55E-01	RP11-622K12.1	0
29	STSS_0055	chr12	6308381	6309381	1001	5	-40.10	4.91E-11	-34.48	8.13E-07	5.62	2.31E-01	-5.31	6.74E-01	CD9	0
30	CGIR_0005	chr7	5647657	5647984	328	8	-34.23	1.02E-09	-27.18	8.03E-07	7.05	1.56E-01	4.14	5.77E-01	FSCN1	1370
31	SEGR_0019	chr3	71478053	71478253	201	4	-29.13	2.52E-04	-30.00	3.45E-07	-0.88	5.59E-01	3.87	5.51E-01	FOXP1	0
32	SEGR_0020	chr3	71478053	71478274	222	4	-28.49	3.35E-04	-25.41	1.28E-05	3.09	4.18E-01	4.93	4.94E-01	FOXP1	0
33	SEGR_0021	chr3	71478171	71478274	104	3	-32.22	1.77E-05	-27.23	4.23E-06	4.99	3.10E-01	5.43	4.24E-01	FOXP1	0
34	SEGR_0040	chr8	42623718	42641148	17431	5	30.19	6.96E-09	28.19	7.57E-10	-2.00	4.44E-01	0.75	7.30E-01	CHRNA6	0
35	SEGR_0087	chr18	63418129	63418179	51	5	32.21	2.55E-06	25.43	5.83E-05	-6.78	1.71E-01	9.54	2.84E-01	CDH7	0

TABLE 17 Panel 2. Caption next page.

TABLE 17 LIST OF FILTERED DIFFERENTIALLY METHYLATED REGIONS. The 35 differentially methylated regions (DMRs) are listed with the following related data: identifier (DMR ID), coordinates (chromosome, start and end positions), length (bp), number of CpG covered, average methylation percentage (%) \pm standard deviation in CSC, non-CSC, iCSC and inon-CSC populations (**panel 1**), differential methylation with q-value for non-CSC vs CSC, non-CSC vs iCSC, CSC vs iCSC, and non-CSC vs inon-CSC pairwise comparisons (%) (**panel 2**), closest gene and distance to the closest gene (bp). A differential methylation > 0 corresponds to hypomethylation (red) and a differential methylation < 0 corresponds to hypermethylation (green). **Chr.**= chromosome; **Diff. Meth.**= differential methylation; **SD**= standard deviation.

KEY POINTS

- The differential methylation analysis of 1,000 bp tiles, 200 bp tiles, regions surrounding transcription start sites (TSSs), CpG island (CGI) regions, and methylation segment regions highlighted 2,596 differentially methylated regions (DMRs) that could undergo methylation changes related to the dedifferentiation of non-CSCs into iCSCs.
- By cross-referencing genes within the 2 kb range of these 2,596 DMRs with genes linked to “stem cell population maintenance” gene ontology, the FANCC, NR2E1, SFRP1, STAG2, STAT3, and WNT7A genes are found associated with at least one hypomethylated DMR (in CSCs compared to non-CSCs) and the BMP7, MED12, PRDM16, STAG2, WNT9B, and ZHX2 genes are found associated with at least one hypermethylated DMR (in CSCs compared to non-CSCs).
- By filtering the 2,596 DMRs based on methylation profiles across subpopulations and minimal number of CpG sites covered in all samples, 35 DMRs associated to 23 unique genes, including the FSCN1, CHRNA6, CDH7, CD9, and PRKAR1B genes, are identified as regions that are likely to undergo methylation changes during dedifferentiation and could participate in the phenotypic switch regulation.

^ [Back to Table of Contents](#)

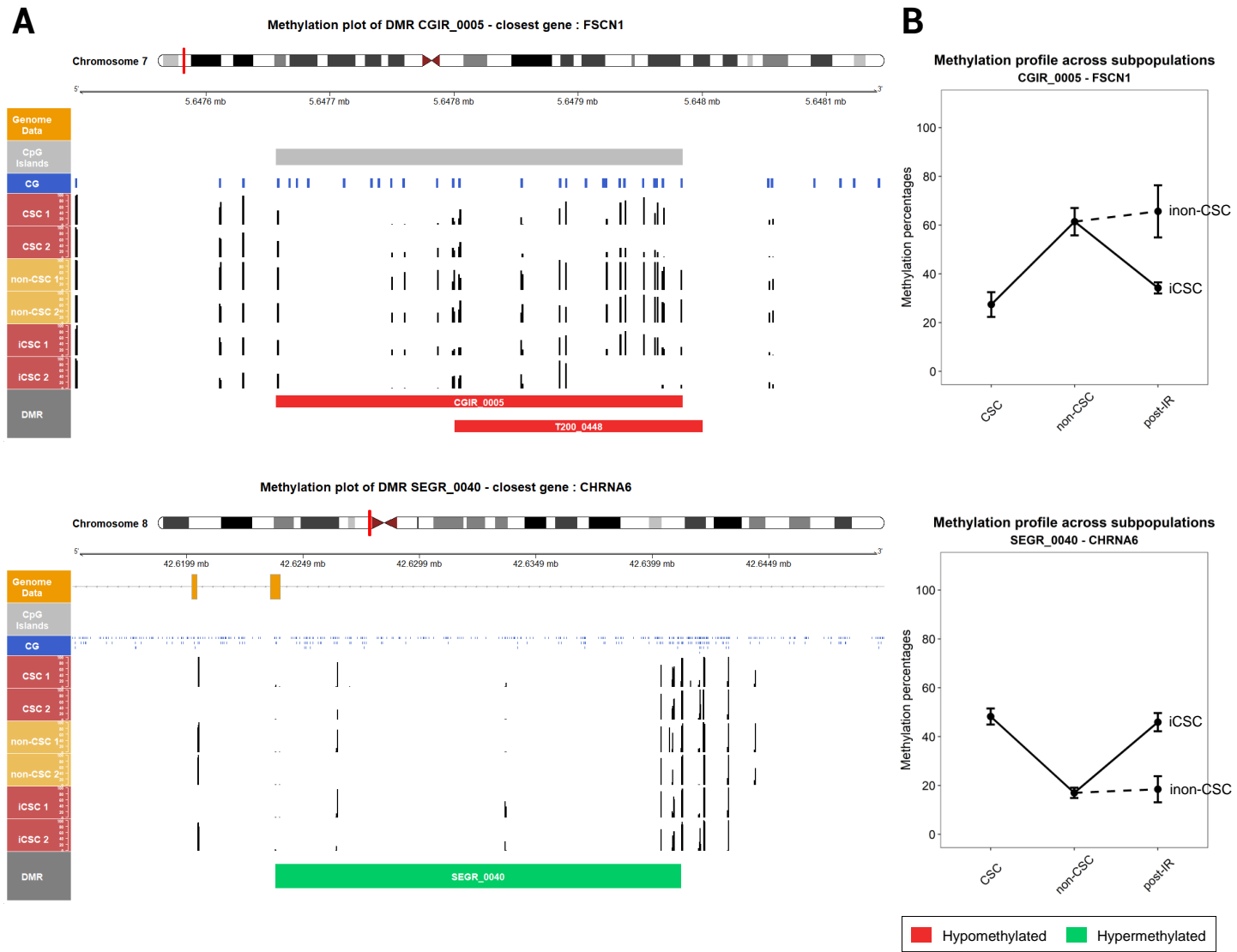


FIGURE 46 *Caption in following pages.*

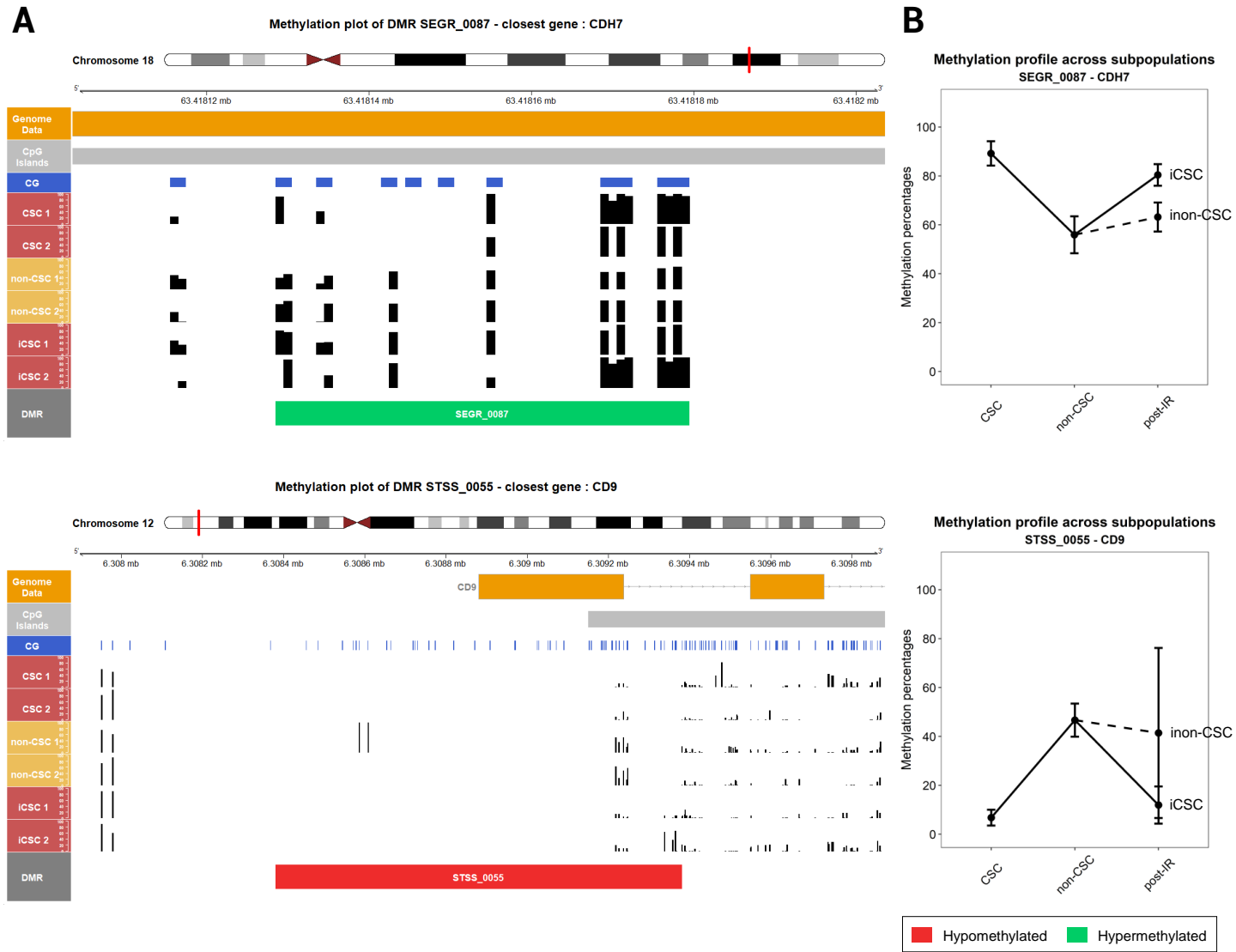


FIGURE 46 Caption next page.

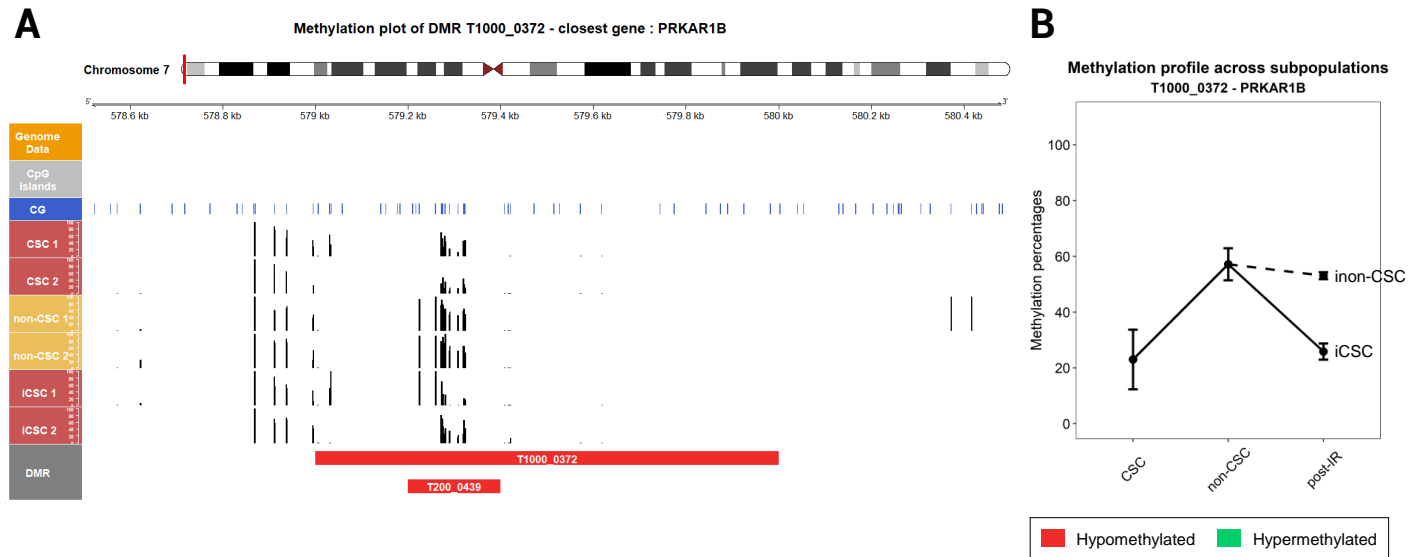




FIGURE 46 METHYLATION PROFILES OF DMRs ASSOCIATED WITH FSCN1, CHRNA6, CDH7, CD9, AND PRKAR1B GENES. **A.** Methylation plots of DMRs close to FSCN1, CHRNA6, CDH7, CD9, and PRKAR1B genes. From top to bottom: chromosome localization; genomic axis; genome data track displaying transcripts along the genome; CpG islands track displaying CGIs along the genome; CG track displaying positions of CpG sites; CSC 1, CSC 2, non-CSC 1, non-CSC 2, iCSC 1 and iCSC 2 tracks displaying CpG methylation percentages histograms; DMR track displaying positions of identified DMRs, hypomethylated ones in CSCs compared to non-CSCs in red, and hypermethylated ones in CSCs compared to non-CSCs in green. **B.** Methylation profiles plots of DMRs close to FSCN1, CHRNA6, CDH7, CD9, and PRKAR1B genes across subpopulations.

2

DEVELOPMENT OF THE ABSP "ANALYSIS OF BISULFITE SEQUENCING PCR" R TOOL FOR REGION-SPECIFIC CPG METHYLATION ANALYSIS



☰ CHAPTER CONTENTS

2.1	 PUBLICATION - "ABSP: AN AUTOMATED R TOOL TO EFFICIENTLY ANALYZE REGION-SPECIFIC CPG METHYLATION FROM BISULFITE SEQUENCING PCR" . . .	209
2.2	ANALYSIS STRATEGY	218
2.2.1	DIRECT-BSP AND CLONING-BSP APPROACHES	218
2.2.2	GLOBAL WORKFLOW	219
2.3	AUTOMATIZATION AND INTERFACE	219
2.3.1	INSTALLATION AND OPENING	220
2.3.2	COMPONENTS OF ABSP	222
2.3.3	APP INTERFACE AND USABILITY	222
2.3.4	SCRIPT AUTOMATIZATION	227
2.3.5	ERROR HANDLING	228
2.3.6	MULTIPLE ANALYSES	231
	 KEY POINTS	235

[^ Back to Table of Contents](#)

2.1  PUBLICATION - "ABSP: AN AUTOMATED R TOOL TO EFFICIENTLY ANALYZE REGION-SPECIFIC CPG METHYLATION FROM BISULFITE SEQUENCING PCR"

Subject Section

ABSP: an automated R tool to efficiently analyze region-specific CpG methylation from bisulfite sequencing PCR

Marie Denoulet^{1,2}, Mathilde Brulé^{1,2}, François Anquez³, Audrey Vincent¹, Julie Schnipper⁴, Eric Adriaenssens¹, Robert-Alain Toillon¹, Xuefen Le Bourhis¹ and Chann Lagadec^{1,2,*}

¹ Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277 - CANTHER - Cancer Heterogeneity Plasticity and Resistance to Therapies, F-59000 Lille, France, ² Institut pour la Recherche sur le Cancer de Lille (IRCL), 59000 Lille, France, ³ CNRS, UMR 8523 - PhLAM - Physique des Lasers Atomes et Molécules, University of Lille, 59000 Lille, France, and ⁴ Laboratory of Cellular and Molecular Physiology, UR UPJV 4667, University of Picardie Jules Verne, Amiens, France.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Nowadays, epigenetic gene regulations are studied in each part of the biology, from embryonic development to diseases such as cancers and neurodegenerative disorders. Currently, to quantify and compare CpG methylation levels of a specific region of interest, the most accessible technique is the Bisulfite Sequencing PCR (BSP). However, no existing user-friendly tool is able to analyze data from all approaches of BSP. Therefore, the most convenient way to process results from the direct sequencing of PCR products (direct-BSP) is to manually analyze the chromatogram traces, which is a repetitive and prone to error task.

Results: Here, we implement a new R-based tool, called ABSP for Analysis of Bisulfite Sequencing PCR, providing a complete analytic process of both direct-BSP and cloning-BSP data. It uses the raw sequencing trace files (*.ab1*) as input to compute and compare CpG methylation percentages. It is fully automated and includes a user-friendly interface as a built-in R shiny app, quality control steps, and generates publication-ready graphics.

Availability and implementation: The ABSP tool and associated data are available on GitHub at <https://github.com/ABSP-methylation-tool/ABSP>.

Contact: chann.lagadec@inserm.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Aside from transcription factor regulations, gene expression can also be activated or repressed by epigenetic modifications directly on nucleotides (DNA methylation) or histones (methylation, acetylation...). In vertebrates, epigenetic regulation is essential to regulate genomic imprinting, X chromosome inactivation, development regulation, cell differentiation, and genome integrity preservation. DNA methylation can affect cytosine and adenine but mostly occurs on a cytosine followed

by a guanine (CpG site). The effect of these modifications on gene transcription has been observed when several grouped CpG within a DNA region, so-called CpG islands, are modified altogether (Greenberg and Bourhis, 2019; Jones, 2012). Specific enzymes, DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B), transfer a methyl group (CH₃) from S-Adenosyl methionine (SAM) on the C5 position of the pyrimidine ring, converting cytosine (C) into 5-methylcytosine (5mC).

Among other methods, the Bisulfite Sequencing PCR (BSP) is the most accessible and conventional method to evaluate methylation levels at single CpG resolution in a mix of DNA molecules (Clark *et al.*, 1994;

Frommer *et al.*, 1992). Even if broad methods have been developed to study DNA methylation, the BSP technique has the benefit of a great sensitivity at a very low cost, compared to other methods using Next Generation Sequencing (NGS) technologies, more sensitive but costly. The BSP assay is thereby the most suited one to quantify DNA methylation of a specific region when large-scale NGS methods are not necessary, especially to get rapid preliminary results, or to validate methylation data from screening experiments such as Reduced Representation Bisulfite Sequencing (RRBS) at specific loci (Chen *et al.*, 2022; Dehdari *et al.*, 2022; Pajares *et al.*, 2021).

DNA methylation estimation methods using bisulfite conversion are based on the selective deamination of cytosine residues by sodium bisulfite treatment, transforming cytosines into uracils whereas 5-methylcytosines are not affected and remain cytosines (Frommer *et al.*, 1992; Hayatsu *et al.*, 1970). Subsequently, the Polymerase Chain Reaction (PCR) regenerates thymines instead of unmethylated cytosines, as both are complementary to adenines, while 5-methylcytosines remain cytosines. Therefore, the original methylated cytosines are distinguishable from the unmethylated ones through Sanger sequencing.

Two approaches to BSP have been described in the literature: direct-BSP and cloning-BSP (Chatterjee *et al.*, 2017). The direct-BSP approach consists in sequencing PCR products directly after PCR amplification of bisulfite-converted DNA. As a mix of DNA molecules with different CpG methylation statuses is being sequenced simultaneously, the quantification of CpG methylation can be assessed in the same way as the quantification of a Single Nucleotide Polymorphism (SNP) (Qiu *et al.*, 2003). Thereby, from the chromatogram trace file, the peak heights ratio of cytosine and thymine signals are used to determine the proportion of methylated cytosines compared to unmethylated ones at CpG sites (Fig. 1) (Jiang *et al.*, 2010; Lewin *et al.*, 2004; Parrish *et al.*, 2012).

In the cloning-BSP approach, PCR products are cloned in vectors and used to transform bacteria. Amplified vectors from individual colonies are extracted for sequencing. Hence, the sequencing of a unique clone, reveals the methylation status of each CpG site of a single PCR product (Li and Tollefsbol, 2011). In the literature, about 10 clones are usually sequenced to get an estimation of the CpG methylation levels of a DNA population with 10% to 20% accuracy (Chen *et al.*, 2022; Li and Tollefsbol, 2011).

In terms of analysis, tools have been developed to analyze cloning-BSP results, exclusively relying on the base-called sequence from the sequencing. For instance, MethTools 2.0, BiQ Analyzer, QUMA (Quantification tool for Methylation Analysis), and BISMA (Bisulfite Sequencing DNA Methylation Analysis) can be cited (Bock *et al.*, 2005; Grunau *et al.*, 2000; Kumaki *et al.*, 2008; Rohde *et al.*, 2010). These tools have been designed to process cloning-BSP data and cannot analyze direct-BSP results as they were not conceived to use the four-dye signal intensity values from chromatograms as an input to interpret the results. Indeed, they determine the methylation statuses of CpG sites of each clone and then calculate the ratio between methylated and unmethylated clones to estimate CpG methylation proportions in the biological sample.

The cloning-BSP approach is mostly used since the direct-BSP one is generally considered less quantitative, due to differences in labeled terminator nucleotides (ddNTPs) incorporation efficiencies and differences in signal relative intensities between the four dyes (Chhibber and Schroeder, 2008; Mikeska *et al.*, 2010). Yet, studies claim that 10 clones are not sufficient to obtain a statistically significant estimation of DNA methylation levels and prone to the direct sequencing of PCR products (Mühlisch *et al.*, 2007; Paul and Clark, 1996; Voss *et al.*, 1998). Besides, direct-BSP is efficient and avoids the multiplication of subclones sequencing costs; it is therefore particularly useful for methylation quantification studies with many samples such as cohorts, or for validation of potential targets identified through screening experiments (Moschny *et al.*, 2020; Schiele *et al.*, 2021).

In the context of The Human Epigenome Project by the Human Epigenome Consortium (2003), the direct-BSP approach was selected to map the CpG methylation levels along the genome for high throughput and cost-effectiveness reasons. Consequently, in 2004, Lewin *et al.* developed an algorithm called ESME (Epigenetic Sequencing Methylation analysis software), to estimate methylation levels from the four-dye chromatogram trace files. However, the software is not up-to-date with the current BSP technology and suffers from accessibility issues as its installation and operation require qualified expertise in a Linux operating system (Akika *et al.*, 2017). So, nowadays, the most convenient way to analyze direct-BSP data still consists in manually retrieving the peak heights to compute methylation percentages of CpG sites, which is time-consuming (dependant on the number of samples and CpG sites per sample), repetitive, prone to errors, and does not include valuable quality control over sequencing data (Jiang *et al.*, 2010; Martisova *et al.*, 2021; Parrish *et al.*, 2012).

Additionally, a step further is required for better visualization and comparison of methylation differences. Once methylation levels are obtained, some graphical visualization of methylation data can be generated, by using a web-based tool called Methylation plotter for example, as well as comparative statistics (Mallona *et al.*, 2014).

Existing tools are not sufficient to provide a full analytic process of BSP results, especially for direct-BSP experiments, in the context of preliminary data or large studies for which the cloning is not appropriate. As it is relevant to sequence the PCR products to estimate methylation percentages before committing to the cloning step, the choice was to apply the same method for both direct-BSP and cloning-BSP results to ensure continuity in the analytic process. By using our new tool ABSP, both approaches of BSP can be analyzed to generate methylation visualization plots and comparative statistics, in an automated and controlled manner, from the Applied Biosystems, Inc. Format (ABI) sequencing files (*ab1*).

2 Approach

For direct-BSP, ratios of the peak heights of the two co-existing C and T signals at CpG positions are used to evaluate the proportion of methylated cytosines (Fig. 1) (Jiang *et al.*, 2010; Lewin *et al.*, 2004; Parrish *et al.*, 2012). The same method can be applied to analyze the subclones sequencings: as the ratio of signal peak heights can either be around 0% or 100%, its calculation reveals the CpG methylation status of individual DNA molecules. Therefore, PCR replicates or clone analysis can give statistical meanings of the degree of methylation among the samples (Fig. 1).

To fully analyze the BSP experiments, 2 main steps are required (Fig. 2). First, the CpG methylation levels of each sample have to be estimated using replicates or clones. In our ABSP-developed process, this step is called *individual analysis*. Next, the *grouped analysis* can be run to compare methylation levels between groups and to find methylation differences.

As presented in Figure 2, each sequencing run of a unique PCR product or a unique clone vector is defined by a combination of information used to track, group, and compare the sample methylation data: (1) the sequence amplified by PCR (unique primer pair used for the BSP experiment), (2) the collection, which describes a separation of samples above groups, it means that samples from different collections cannot be compared (e.g. cell lines or organs), (3) the group, which is the experimental condition to compare (e.g. treatment a or b), and (4) the replicate number for direct-BSP (repetition identifier) or the clone number for cloning-BSP (clone identifier).

Additionally, sequencing reads from both directions can be provided for each unique DNA sample, using a forward and reverse primer, to

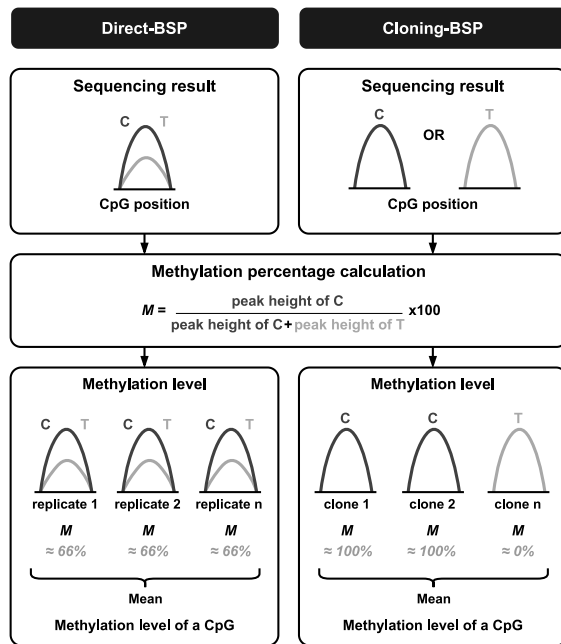


Fig. 1. Analysis strategy differences between the two Bisulfite Sequencing PCR approaches, direct-BSP and cloning-BSP. As the direct-BSP method consists in sequencing the mix of PCR products, the methylation calculation based on chromatogram C/T peak heights gives directly an estimation of the CpG methylation level, that can be replicated for statistical significance determination. In the cloning-BSP method, as the CpG methylation status in each clone is revealed by the chromatogram peak heights as well, the sequencing of several clones gives an estimation of the CpG methylation proportion.

maximize the sequence coverage and increase data robustness as both sequencing reads can overlap.

3 Methods

3.1 General structure

Each one of the two main parts of ABSP, the *individual* and *grouped* analyses, corresponds to an R markdown script (using the *markdown* R package), thereby generating two different types of analysis reports, one specific for individual sample results and the other for grouped samples analysis results. These two analyses can be launched through a shiny app, in which the *individual analysis* tab and the *grouped analysis* tab serve to enter the input parameters, required for report rendering (using the *knitr* R package). Once the inputs are filled and the analysis is launched, the corresponding script processes the analysis, exports several output files, and produces the analysis report as an HTML file (*.html* extension), summarizing all the results and serving as a record of them. An additional tab called *multiple analyses* serves to launch several analyses, *individual* ones, *grouped* ones, or both, in one click, using filled tables (*.xlsx* or *.csv* files) as input entries.

3.2 Individual analysis

The *individual analysis* aims to compute the CpG methylation percentages from the chromatogram trace files of each individual sample at each CpG site, using the signal peak height values.

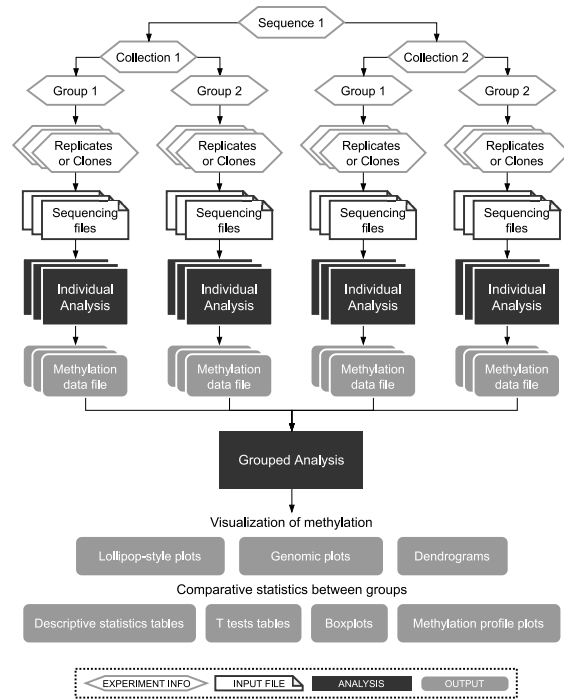


Fig. 2. General workflow of ABSP. The analysis is divided into two main steps: the *individual analysis* and *grouped analysis*. The first one serves to control the sequencing quality and compute methylation levels for each individual sample, whereas the second one gathers all sample methylation results to generate visualization plots and process comparative statistics between groups.

Input required

Three inputs are required to proceed through the analysis: (1) the sample combination of information, to affiliate the methylation results to the correct sample (Fig. 2), (2) the genomic reference sequence, its genomic coordinates, and the strand amplified during PCR (as the bisulfite converts cytosines into uracils, the two DNA strands are no longer complementary, only one can be amplified with a unique set of primers), which have to be provided in a FASTA file (*.fasta* extension), and (3) the chromatogram trace files in ABIF format (*.ab1* extension) of the sequencing reads in forward and reverse directions. In Figure 3, the sequencing results are numbered #1 and #2, as the direction must not be specified and will be automatically determined during the analysis.

Processing the reference DNA

First, the CpG positions are retrieved from the reference DNA, and their coordinates are found by correlating positions and reference coordinates (Fig. 3, *Find CpG coordinates* frame). Matches for CG dinucleotides in both plus and minus strands give the start and end positions of each CpG. CpG coordinates are calculated using the start and end coordinates of the reference sequence and CpG positions on the reference sequence (e.g. CpG site at positions 100-101 on plus strand: cytosine coordinate = seq_start + position - 1 = 6,000 + 100 - 1 = 6,099).

In Figure 3, the second line of process panels represents BSP experiment steps reproduced *in silico* during the analysis. The theoretical bisulfite conversion of the reference DNA is realized using the amplified strand sequence (Fig. 3, *Bisulfite Conversion* frame). As the PCR regenerates the opposite strand of the DNA template, both sequences are retrieved (Fig. 3, *PCR Amplification* frame): (1) sequence from the

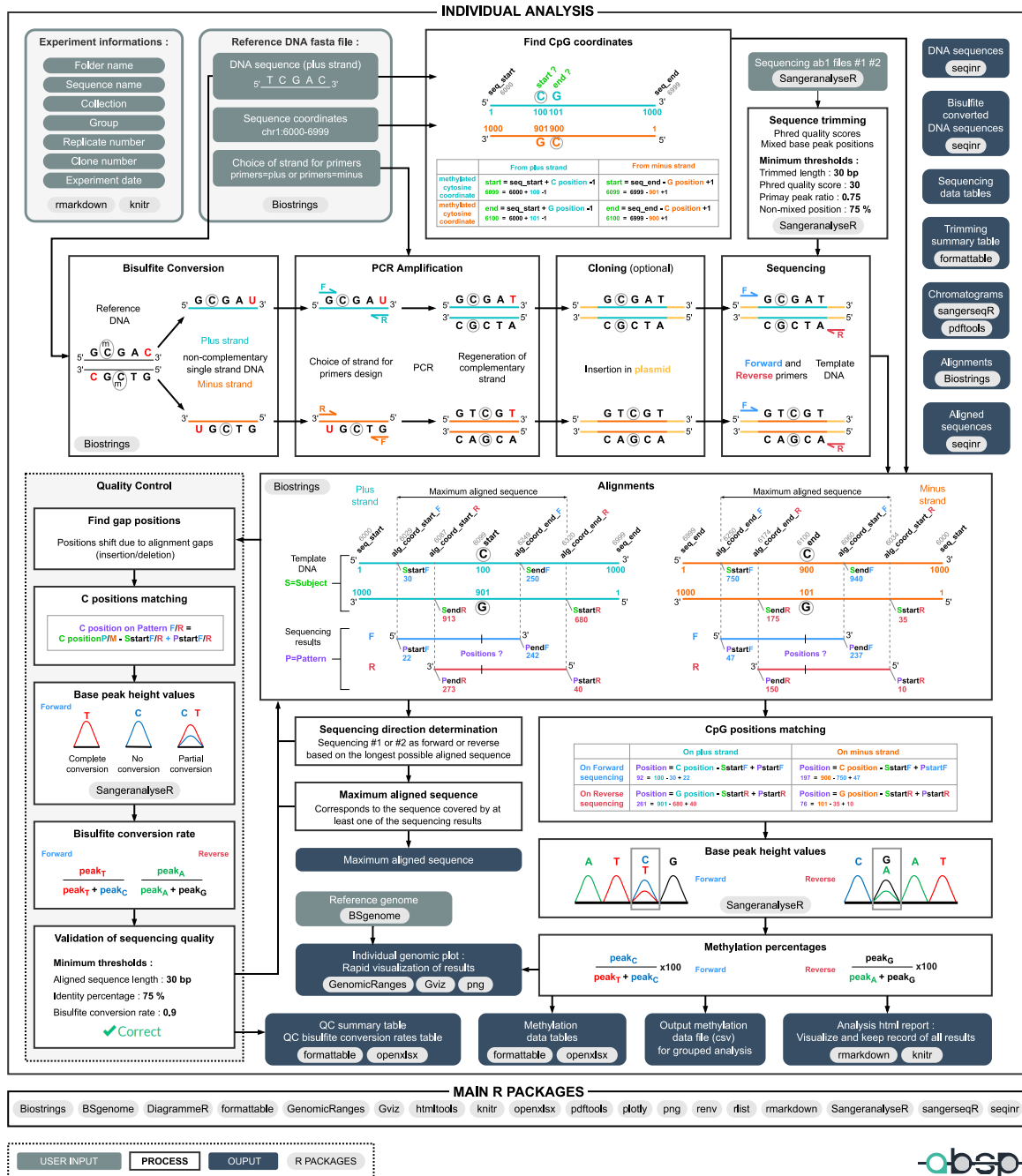


Fig. 3. Detailed workflow of the individual analysis step. To illustrate the process, an arbitrary reference sequence of 1000 bp was chosen, with genomic coordinates between 6,000 and 6,999 and a CpG site at positions 100-101 on the plus strand. Other numbers, such as aligned sequences borders, were also arbitrarily chosen for example purposes.

amplified strand, or sense strand as it will serve as the template for the sequencing in the forward direction (upper strand in Fig. 3), and (2) the sequence from the opposite strand, or antisense strand as it will serve as the template for the sequencing in the reverse direction (lower strand in Fig. 3) (Fig. 3, Sequencing frame).

Trimming of sequencing results

As the extremities of the sequencing reads are prone to off-scale signals and errors, these inaccurate parts must be removed. To determine the correct positions where the sequencing read should be trimmed, 2 parameters are

used: the base calling error probability and the signal peak height (Fig. 3, *Sequencing trimming* frame).

The first trimming method is based on the base calling Phred quality scores, to remove parts susceptible to having base calling errors. This step is performed by the *sangeranalyseR* R package, using the modified Mott's trimming algorithm (M1 method) with a base calling error probability (P) default cutoff of 0.001%, equivalent to a Phred quality score (Q) of 30 ($Q = -10 \times \log_{10}(P)$) (Chao *et al.*, 2021).

Based on the peak height values from the chromatogram, the second way of trimming aims to remove extremities where signals are mixed. For each base, the ratio of the primary peak height signal over the total of all signals peak height is calculated using the following formula: if $peak_C > \{peak_A, peak_T, peak_G\}$, primary peak ratio = $peak_C / (peak_A + peak_T + peak_G + peak_C)$. A base position is considered as "non-mixed" if the primary peak ratio is above the threshold, set by default to 0.75. Then, all the possible trimmed sequences are obtained by selecting the longest sequence for which the boundaries are of n (n from 3 to 15) consecutive non-mixed positions. Among those trimmed sequences, the one with a percentage of non-mixed positions above the threshold (default is 75%) with the minimum of consecutive non-mixed positions at boundaries is kept.

Finally, the overlap between both sequences, from the quality score trimming and the mixed base peak trimming, gives the final trimmed sequence used for the following steps. If one of the trimming methods fails or if the final trimmed sequence parameters are below thresholds (length, average Phred score, and percentage of non-mixed positions, Fig. 3, *Sequencing trimming* frame), the sequencing will not be used to compute methylation percentages.

Alignments of trimmed sequencing reads with template DNA sequences

To correlate nucleotide positions on the sequencing reads with CpG positions on the template DNA, the alignment of sequences is performed (local pair-wise alignment). In cloning-BSP experiments, sequencing primers are often chosen on the vector backbone. As PCR products can be inserted in either direction, it is crucial to determine the direction of the sequencing within the analytic process. Trimmed sequences are first aligned with both sense and antisense sequences of the converted template DNA. The longest alignment is considered the correct template (Fig. 3, *Alignments* frame).

Knowing the positions of the first nucleotide on template DNA (Subject, S) and sequencing read (Pattern, P), respectively $SstartF/PstartF$ for forward sequencing, and $SstartR/PstartR$ for the reverse sequencing, a direct correlation is used to find cytosine positions on the trimmed sequencing results. So, as an example, on the forward strand, if the cytosine is at the position 100 on the template (S), the $SstartF = 30$ and the $PstartF = 22$, cytosine position = cytosine position on template - $SstartF + PstartF = 100 - 30 + 22 = 92$.

The maximum aligned sequence corresponds to the sequence covered by at least one of the sequencing reads, and its coordinates are determined by the correlation of genomic coordinates and aligned positions (alg_coord_start and alg_coord_end).

Quality control of the aligned sequencing results

To check the concordance between the template DNA and the sequencing results, the aligned sequences are controlled through several steps: (1) gap positions determination, (2) C positions matching for bisulfite conversion rate calculation, (3) retrieval of peak height values for each C position outside CpG sites, (4) bisulfite conversion rate calculation, and (5) validation of the sequencing quality.

As the retrieval of peak height values for methylation calculation is based on the start positions of aligned sequences, the presence of a gap, insertion, or deletion in either the template DNA or the sequencing result, causes a position shift that needs to be corrected for the CpG position matching step. The most important criteria to validate the quality of a sequencing result is the bisulfite conversion efficacy. To assess its efficacy, the bisulfite conversion rate is computed for each cytosine position outside CpG sites and the average rate on the sequence must be higher than the provided threshold (default is 0.9). First, C positions have to be retrieved based on the alignments, with the same method as explained above, by matching the positions of aligned sequences. Then, the peak height values of each base at these positions are used to calculate the bisulfite conversion rates with the following formula, for the forward sequence: bisulfite conversion rate = $peak_T / (peak_C + peak_T)$ and for the reverse sequence: bisulfite conversion rate = $peak_A / (peak_C + peak_A)$. Finally, the alignments and quality control steps provide the aligned sequence length, identity percentage, mismatches positions, insertion/deletion positions, and the average bisulfite conversion rate (Fig. 4B). For a sequencing result to be considered as correct, the length, identity percentage, and average bisulfite conversion rate have to be higher than the defined thresholds, set to 30 bp, 75%, and 0.9 respectively by default.

Methylation calculation

The methylation percentage of each CpG site is calculated using the peak height values corresponding to the intensity of the dye signal, with the following formula: methylation percentage = $peak_C / (peak_C + peak_T) \times 100$ or = $peak_C / (peak_C + peak_A) \times 100$ for the forward and reverse sequencing results respectively (Figs. 4C and D).

Outputs

The main output result of the *individual analysis* is the methylation data table, used as input for the *grouped analysis* afterward. To visualize the methylation levels of the analyzed sample, a plot displaying the genomic sequence, the CpG positions, and the methylation levels as a grey gradient is produced (Supplementary Fig. S1). This genomic plot can serve as a control of the coordinates, as CpG site coordinates must match the sequence colors of CG dinucleotides.

3.3 Grouped analysis

Methylation data from individual analysis

As input, the methylation data files saved by the previous *individual analyses* are automatically retrieved based on the selected folder and sequence name. Methylation data from all samples are processed and gathered.

For each individual clone, the methylation percentages found based on signal peak ratios are converted into methylation statuses. By default, for each CpG, a methylation level between 0% and 20% is considered as an unmethylated status and a methylation level between 80% and 100% as a methylated one. Partial methylation, between 20% and 80%, is considered defective and is annotated as not available. For one clone, if the number of partially methylated CpGs is important (above 20% by default) the clone is considered as a potential mix of clones and therefore all of its methylation data is annotated as not available.

Generation of plots to visualize methylation levels

To generate visualization plots, several plot parameters are required: (1) the label type for CpG positions (CpG coordinates, CpG numbers, or none), (2) the collection separation, whether or not samples from different collections have to be displayed on the same plot, (3) the group order for display, and (4) the sample ordering on the ordinate axis (as it is, by groups, by methylation levels, or by clusters).

To represent CpG methylation levels, the lollipop-style plots are largely used in the literature. They illustrate CpG levels as circles with methylation levels either as a black and white scale for clone methylation status or as a grey gradient for methylation level (Figs. 4E and G). Most plots generated by ABSP were built using the functions of the Methylation plotter tool as a reference (Mallona *et al.*, 2014).

As for the *individual analysis*, methylation levels are also pictured by genomic plots, displaying the genomic sequence, CpG positions, and CpG methylation of samples as a grey-scale heatmap along the sequence (Supplementary Fig. S2).

Comparative statistics

As the purpose of the BSP experiment is to compare results from different conditions over methylation levels, several outputs are generated: tables with the two-by-two comparisons of groups with Student's T-test *p*-values, boxplots representing the methylation means of each CpG, and boxplots with the means of all CpG analyzed gathered with Student's T-test *p*-values as well (Figs. 4H and I, Supplementary Fig. S3), and finally methylation profile plots displaying the methylation levels as line plots along the sequence with Kruskal-Wallis *p*-values per CpG, to identify the sites with significant differences among the groups (Fig. 4F) (Mallona *et al.*, 2014).

4 Application

Both high-methylated and low-methylated human genomic DNA (80-8061-HGHM5 and 80-8062-HGUM5 from EpigenDx) were treated with sodium bisulfite and cleaned up. An upstream promoter region of the CDH1 gene, covering 17 CpG sites was amplified through a touchdown PCR protocol using specific primers, 5' tailed with standard primers T3 or BGH Reverse. The 259 bp long amplicons were directly sequenced in both directions, in triplicates to allow statistical analysis (for additional details on the method, see the Supplementary Materials).

Sequencing results were processed and analyzed using the ABSP workflow described for direct-BSP analysis. Essential results from the *individual analysis* and *grouped analysis* reports are respectively displayed in Figure 4 top and bottom panels.

The CG #8 from the high-methylated DNA #3 sample is displayed to illustrate the analysis process and outputs (Fig. 4). After alignment with the reference sequence and validation of both sequencing results through quality control (Fig. 4B), the peak height values corresponding to each base at the CG #8 cytosines positions are retrieved (Fig. 4C). The C and T peak heights are used to compute the methylation percentage from the forward sequencing, and the A and G peak heights from the reverse sequencing, as displayed in Figure 4C. After combining methylation results from both sequencing reads, the average methylation percentage and standard deviation are computed and these data will be used in the *grouped analysis* (Fig. 4D). For the CG #8 illustrated in Figure 4, the sequencing analysis reveals methylation of 71.13% for the forward result and 82.65% for the reverse, given an average methylation of 76.89% (Figs. 4C and D). As CG position numbers are determined based on the reference DNA in the *individual analysis* and are then reset in the *grouped analysis*, the previously described CG #8 corresponds now to the CpG site #4 covered by at least one of the sequencing.

As the *grouped analysis* aims to facilitate the interpretation of methylation data from all samples, several graphics are generated. First, in the lollipop-style plot displaying methylation of all samples, the difference between the low-methylated and high-methylated samples is clearly visible thanks to the grey scale (Fig. 4E). In addition, missing points, inconsistent methylation levels between replicates or clones, and methylation patterns can be easily found on this type of plot. In the high-methylated DNA #3 sample, it is particularly noticeable that the CpG site #4 has a slightly

lower methylation level (76.89%) compared to the 2 other high-methylated DNA replicates (Fig. 4E, green circle). For unknown reasons, the forward sequencing reads were not clean enough and failed to pass the trimming and/or quality control steps for 5 out of 6 samples, explaining the missing data points, covered neither by the forward sequencing read nor by the beginning of the reverse sequencing read. For a robust comparison of methylation between groups, the methylation profile plot indicates the CpG sites for which the difference in methylation level is significant among groups, which is the case here for all the CpG covered in the 2 groups (Fig. 4F). Additionally, the lollipop-style plot displaying the methylation means of groups provides less information but gives an efficient overview of methylation differences between groups (Fig. 4G).

To complement the comparative analysis, boxplots of each CpG site and the boxplot of means of CpG methylation, indicate the distribution of methylation among the groups as well as the significance of methylation differences between groups two-by-two (Figs. 4H and I). The CpG site #4 has a methylation percentage of 0.47% ($\pm 0.81\%$) in low-methylated DNA and 81.44% ($\pm 3.96\%$) in high-methylated DNA, with a statistical *p*-value of 0.00051. Also, the mean methylation rate of the sequence CpG is 1.43% ($\pm 1.87\%$) in low-methylated DNA and 95.54% ($\pm 0.67\%$) in high-methylated DNA, with a statistical *p*-value of $2.34e^{-05}$, confirming the difference of methylation of the analyzed sequences. All the data associated with this example (inputs, reports, outputs) are provided along with the ABSP files, available at <https://github.com/ABSP-methylation-tool/ABSP>.

5 Discussion

For this work, we developed a modern and useful tool to analyze both direct and cloning approaches of BSP. As ESME software is the reference for such studies, we compared results obtained from ESME to ABSP, and found several differences. First, ESME performs a normalization of cytosines signals as it assumes that the less frequent base signals are overscaled by the basecaller (example of ESME results in Supplementary Fig. S4) (Lewin *et al.*, 2004). However, ESME was developed in the early stages of the BSP technology, and nowadays basecallers have been improved and do not exaggerate the missing base, its normalization step is therefore no longer required and may introduce biases in methylation percentages calculated (Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems, 2007, https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_039258.pdf). Hence, ABSP does not apply any changes to the peak height values retrieved from chromatogram trace data, as performed in other studies (Jiang *et al.*, 2010; Parrish *et al.*, 2012) (comparison of ESME and ABSP results in Supplementary Tab. S1).

Also, ABSP provides several key advantages compared to ESME. The main added values of ABSP is a built-in comparative analysis step, including methylation data visualization, with ready-to-publish graphics, and statistical tests, to help researchers answer the experimental hypothesis. Moreover, by being able to process both direct-BSP and cloning-BSP data, ABSP provides an analysis continuity, from preliminary data by direct-BSP up to validation by cloning-BSP. In terms of accessibility, as only R and RStudio are required, ABSP can operate on every operating system supporting both software (Windows, Linux, macOS). More importantly, the full automation of the analysis and the user-friendly interface makes ABSP accessible to users without expertise in R, such as most of biologists.

Yet, ABSP is still an adaptable tool for R accustomed users, as the code can be modified to be adapted to experiments or user needs. For example, minor modifications such as threshold adjustments or plot customization are possible (help to implement the modifications is included in the user



Fig. 4. Output results from ABSP analysis. (A) Portions of the trimmed chromatogram traces from forward and reverse sequencing reads of the CDH1 sequence in the high-methylated DNA sample, replicate #3. The CG #8 is highlighted and corresponds to the 8th CG dinucleotide of the reference sequence. (B) Table of quality control summary after alignments of trimmed sequencing results with the reference sequence. (C) Tables of the methylation calculation per CpG using peak height values from both chromatogram traces. (D) Table of combined methylation data from both sequencing results, with the methylation average and standard deviation per CpG. (E) Lollipop-style plot of CpG methylation levels on the CDH1 sequence, with all the samples displayed and CpG sites placed proportionally to their coordinates. The CG #8 of the high-methylated DNA #3 sample, detailed in the *individual analysis* panel, is highlighted in green in the lollipop-style plot and corresponds to the 4th CpG site on the covered sequence. The crosses represent unavailable data for CpG sites not covered by the trimmed sequencing results in some samples. (F) Methylation profile plot displaying average methylation levels of CpG along the CDH1 sequence in each group as a line plot. Symbols represent significance levels of Kruskal-Wallis test *p*-values (* *p* ≤ 0.05, ** *p* ≤ 0.01, *** *p* ≤ 0.001, **** *p* ≤ 0.0001). (G) Lollipop-style plot of average CpG methylation levels on the CDH1 sequence in each group. (H) Boxplot of the CpG #4 average methylation levels in each group. (I) Boxplot of the methylation percentage means of all CpG positions on the CDH1 sequence in each group. In boxplots, symbols represent significance levels of Student's T-test *p*-values (* *p* ≤ 0.05, ** *p* ≤ 0.01, *** *p* ≤ 0.001, **** *p* ≤ 0.0001).

guide). Other major modifications such as expanding the analysis to include CHG and CHH sites (H being non-Gs bases), can be developed and implemented to ABSP code in a future version, to fit it to DNA methylation study in plant models. Altogether, ABSP provides a new and easy way to process sequencing data from Bisulfite Sequencing PCR experiments and help researchers to compare methylation profile for a given sequence.

Acknowledgements

We thank Frank Bonardi for his help reviewing the R coding. Marie Denoulet was supported by a grant fellowship from the Ligue Contre le Cancer. This work was also supported by the Institute de Recherche contre la cancer de Lille (IRCL charities).

This project was possible thanks to the work done by developers on the Methylation plotter tool (Mallona *et al.*, 2014) and thanks to several R

packages. All the packages used are cited in the ABSP user guide available for download.

Funding

This research was funded by Institut National du Cancer (INCA), Ligue Nationale contre le cancer (Ligue), Association pour la Recherche contre le Cancer (ARC), [ARC_INCa_LNCC_8068], by the "Ruban Rose" charity [Prix Ruban Rose Avenir 2014], by the Ligue contre le Cancer - Septentrion committee, and by Institut de la Recherche contre le Cancer de Lille (IRCL charities).

References

Akika, R. *et al.* (2017). Region of interest methylation analysis: a comparison of MSP with MS-HRM and direct BSP. *Molecular Biology*

- Reports*, **44**(3), 295–305.
- Bock, C. *et al.* (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**(21), 4067–4068.
- Chao, K.-H. *et al.* (2021). sangeranalyseR: simple and Interactive Processing of Sanger Sequencing Data in R. *Genome Biology and Evolution*, **13**(3), evab028.
- Chatterjee, A. *et al.* (2017). Tools and Strategies for Analysis of Genome-Wide and Gene-Specific DNA Methylation Patterns. *Methods in Molecular Biology*, **1537**, 249–277.
- Chen, R. *et al.* (2022). DNA methylation of miR-138 regulates cell proliferation and EMT in cervical cancer by targeting EZH2. *BMC Cancer*, **22**(1), 488.
- Chhibber, A. and Schroeder, B. G. (2008). Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing. *Analytical Biochemistry*, **377**(1), 46–54.
- Clark, S. J. *et al.* (1994). High sensitivity mapping of methylated cytosines. *Nucleic acids research*, **22**(15), 2990–2997.
- Dehdari, H. *et al.* (2022). CYP1A1 contiguous hypermethylation within a putative CpG block is associated with breast cancer progression: Feasibility to define boundary motives. *Experimental Cell Research*, **413**(1), 113062.
- Frommer, M. *et al.* (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, **89**(5), 1827–1831.
- Greenberg, M. V. C. and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, **20**(10), 590–607.
- Grunau, C. *et al.* (2000). MethTools—a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Research*, **28**(5), 1053–1058.
- Hayatsu, H. *et al.* (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, **9**(14), 2858–2865.
- Jiang, M. *et al.* (2010). Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Laboratory Investigation*, **90**(2), 282–290.
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**(7), 484–492.
- Kumaki, Y. *et al.* (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Research*, **36**(Web Server issue), W170–W175.
- Lewin, J. *et al.* (2004). Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics*, **20**(17), 3005–3012.
- Li, Y. and Tollefsbol, T. O. (2011). DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis. *Methods in Molecular Biology*, **791**, 11–21.
- Mallona, I. *et al.* (2014). Methylation plotter: a web tool for dynamic visualization of DNA methylation data. *Source Code for Biology and Medicine*, **9**(1), 11.
- Martissova, A. *et al.* (2021). DNA Methylation in Solid Tumors: Functions and Methods of Detection. *International Journal of Molecular Sciences*, **22**(8), 4247.
- Mikeska, T. *et al.* (2010). The implications of heterogeneous DNA methylation for the accurate quantification of methylation. *Epigenomics*, **2**(4), 561–573.
- Moschny, N. *et al.* (2020). DNA Methylation of the t-PA Gene Differs Between Various Immune Cell Subtypes Isolated From Depressed Patients Receiving Electroconvulsive Therapy. *Frontiers in Psychiatry*, **11**, 571.
- Mühlich, J. *et al.* (2007). Frequent but borderline methylation of p16INK4a and TIMP3 in medulloblastoma and sPNET revealed by quantitative analyses. *Journal of Neuro-Oncology*, **83**(1), 17–29.
- Pajares, M. J. *et al.* (2021). Methods for analysis of specific DNA methylation status. *Methods*, **187**, 3–12.
- Parrish, R. R. *et al.* (2012). Direct Bisulfite Sequencing for Examination of DNA Methylation with Gene and Nucleotide Resolution from Brain Tissues. *Current Protocols in Neuroscience*, **60**(1), 7.24.1–7.24.12.
- Paul, C. L. and Clark, S. J. (1996). Cytosine Methylation: Quantitation by Automated Genomic Sequencing and GENESCAN Analysis. *BioTechniques*, **21**(7), 126–133.
- Qiu, P. *et al.* (2003). Quantification of single nucleotide polymorphisms by automated DNA sequencing. *Biochemical and biophysical research communications*, **309**(2), 331–338.
- Rohde, C. *et al.* (2010). BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*, **11**(1), 230–230.
- Schiele, M. A. *et al.* (2021). Serotonin Transporter Gene Promoter Hypomethylation as a Predictor of Antidepressant Treatment Response in Major Depression: A Replication Study. *International Journal of Neuropsychopharmacology*, **24**(3), 191–199.
- Voss, K. O. *et al.* (1998). Combating PCR Bias in Bisulfite-Based Cytosine Methylation Analysis. Betaine-Modified Cytosine Deamination PCR. *Analytical Chemistry*, **70**(18), 3818–3823.

2.2 ANALYSIS STRATEGY

The Analysis of Bisulfite Sequencing PCR (ABSP) tool was developed to provide an efficient and accessible way to analyze sequencing data from both direct-BSP and cloning-BSP experiments. The currently available tools are limited, therefore ABSP aims to address this shortcoming, to facilitate the study of CpG methylation levels for researchers.

2.2.1 DIRECT-BSP AND CLONING-BSP APPROACHES

As explained in the publication, direct-BSP and cloning-BSP data differ in terms of CpG methylation information. The CpG methylation level in direct-BSP relies on the ratio of C and T bases fluorescence signal intensities — C and T peak height values — while in cloning-BSP it relies on the ratio of C and T bases from several clones (Figure 47). To use the same scripts to analyze data from both approaches, the methylation calculation method used on direct-BSP data is applied to cloning-BSP data, using the following formula based on peak height values at a unique cytosine position in a CpG context, for forward and reverse sequencing runs respectively:

$$\text{Methylation percentage (forward sequencing)} = \frac{\text{peak height of C}}{\text{peak height of C} + \text{peak height of T}} \times 100$$

$$\text{Methylation percentage (reverse sequencing)} = \frac{\text{peak height of G}}{\text{peak height of G} + \text{peak height of A}} \times 100$$

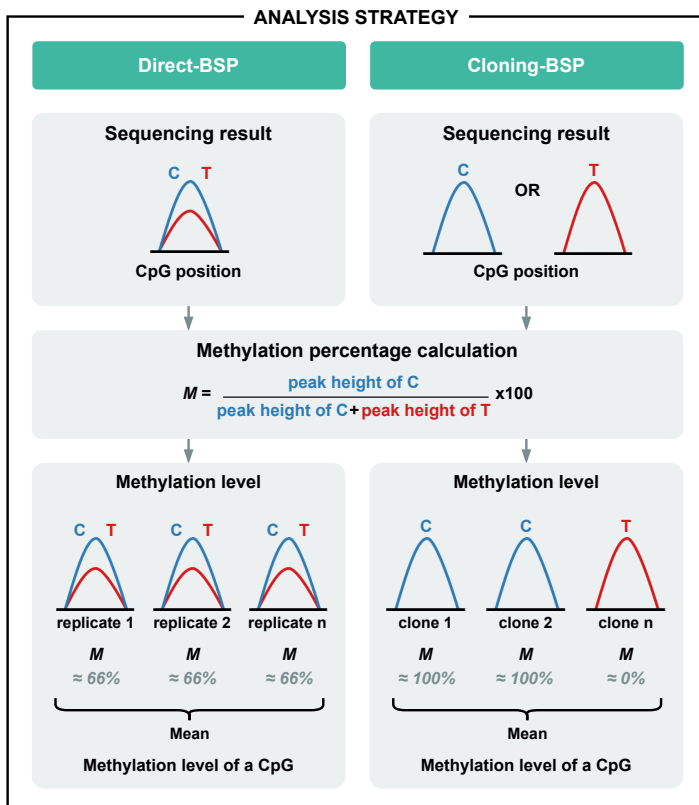


FIGURE 47 ANALYSIS STRATEGY DIFFERENCES BETWEEN THE TWO BISULFITE SEQUENCING PCR APPROACHES, DIRECT-BSP AND CLONING-BSP. As the direct-BSP method consists in sequencing the mix of PCR products, the methylation calculation based on chromatogram C/T peak heights directly gives an estimation of the CpG methylation level, that can be replicated for statistical significance determination. In the cloning-BSP method, as the CpG methylation status in each clone is revealed by the chromatogram peak heights as well, the sequencing of several clones gives an estimation of the CpG methylation proportion.

Indeed, for cloning-BSP data, it results in methylation levels of around either 0% or 100%, revealing respectively the unmethylated or methylated methylation status of the CpG site. Additionally, this method allows for the detection of artifacts or mixed clones when CpG sites are found partially methylated in clones. Then, methylation percentages are still computed using the ratio of methylated clones over the total number of clones (Figure 47).

For direct-BSP, as the signal intensities ratio directly reveals the ratio of methylated cytosines over unmethylated ones, data from only one sequencing run is sufficient to get an estimation of the methylation level. Yet, having several replicates is necessary for the robustness of results, to compute average methylation levels and to run statistical tests (Figure 47).

2.2.2 GLOBAL WORKFLOW

The ABSP analysis is divided into 2 main steps (Figure 48): ① the individual analysis that aims to compute CpG methylation levels in each sample, and ② the grouped analysis that aims to group samples for visualization of methylation data and statistical comparisons between groups of samples to find significant methylation differences.

As illustrated in Figure 48, each sample, corresponding to a unique PCR product or a unique clone vector, is identified by a combination of information:

1. The **sequence** amplified by PCR, from a unique primer pair used for the BSP experiment (e.g. CDH1, CDH1 promoter).
2. The **collection** (optional), which describes a separation of samples above groups, it means that samples from different collections cannot be compared (e.g. cell lines, organs).
3. The **group**, which is the experimental condition to compare (e.g. control, treatment a or b).
4. The **replicate number** for direct-BSP (repetition identifier) or the **clone number** for cloning-BSP (clone identifier).

Lastly, from each sample, the sequencing can be performed in both directions, using a forward and a reverse primer. To maximize the sequence coverage and increase data robustness as both sequencing reads may overlap, the sequencing files from both directions should be provided. However, even if it is not recommended, the analysis can be run with only one sequencing file.

Further practical information about steps to launch analyses, lists of input information, lists of output files, figures of the different plots, folders management, or software license, are all available in the ABSP user guide provided in Appendix 3 “ABSP user guide” at page 338.

2.3 AUTOMATIZATION AND INTERFACE

This section provides additional information on the structure and interface of ABSP enabling its automatization, which is not explained in detail within the publication, as well as portions of R code to highlight the development process.

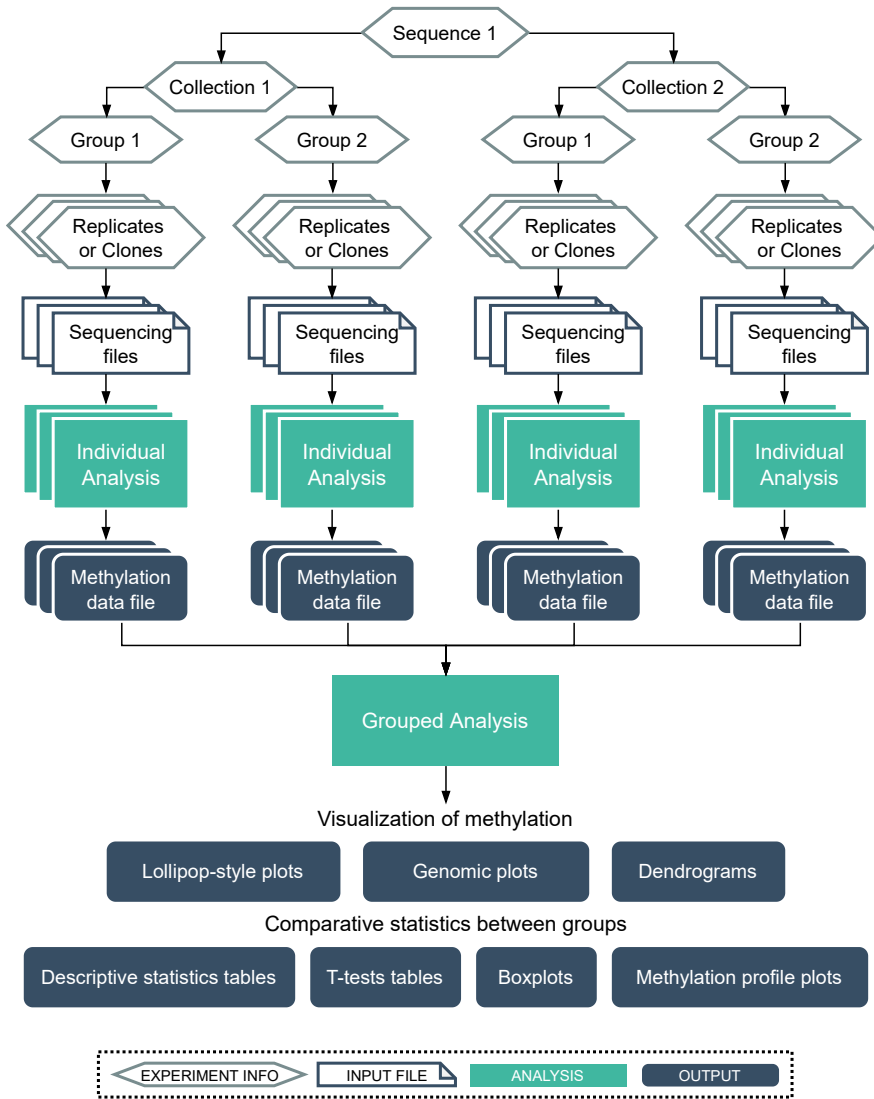


FIGURE 48 GENERAL WORKFLOW OF ABSP. The analysis is divided into two main steps: the ① individual analysis and ② grouped analysis. The first one serves to control the sequencing quality and compute methylation levels for each individual sample, whereas the second one gathers all sample methylation results to generate visualization plots and process comparative statistics between groups.

2.3.1 INSTALLATION AND OPENING

Files of ABSP are available to download on a GitHub repository at <https://github.com/ABSP-methylation-tool/ABSP>. As ABSP uses the R programming language, its utilization requires the R software and the RStudio desktop application, which is a development environment for R. Thus, ABSP can be run on every operating system supporting both R and Rstudio.

To open ABSP, the user must open the “*ABSP Rproject.Rproj*” project file in RStudio, followed by the “*app.R*” file within the ABSP R project in RStudio. Then the “Run App” button can be pressed to open the app interface, either in an RStudio window or in the default web browser.

To get the required R packages for ABSP to run, a package installation step is directly provided at the beginning of the app code to automatically install and load packages within the project. Therefore, the first time the app is opened takes longer as packages need to be installed.

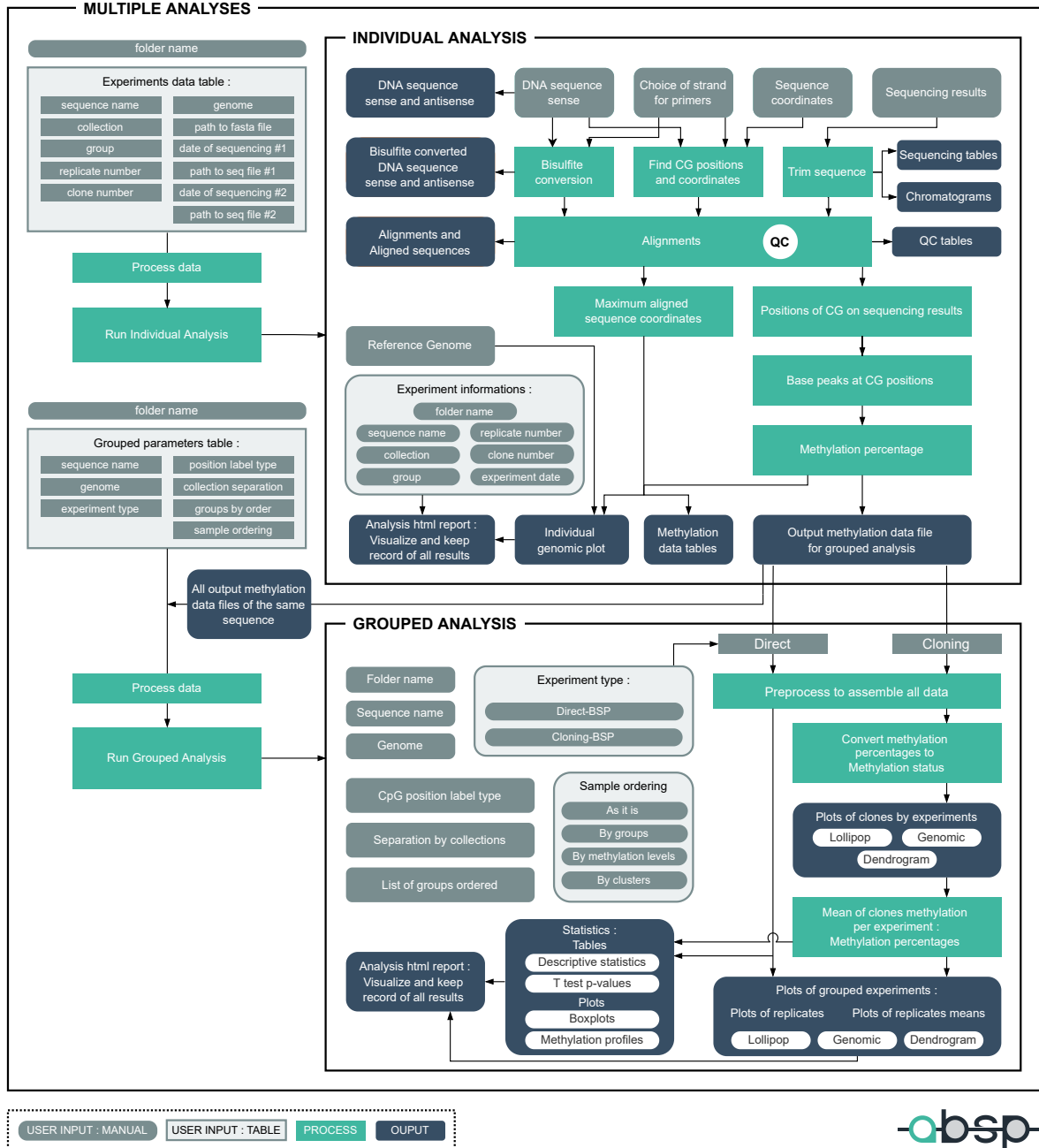


FIGURE 49 DETAILED WORKFLOW OF ABSP. The two main scripts of the analysis, the individual analysis and grouped analysis, can be launched by the manual entry of parameters. To launch multiple analyses at once, inputs provided as tables can be used to automatically launch both individual and grouped analyses.

2.3.2 COMPONENTS OF ABSP

The 2 main steps of ABSP, individual and grouped analyses, are processed by 2 distinct R markdown (RMD) scripts. An RMD script is written in the markdown language (an easy-to-write plain text format) and contains chunks of embedded R code, using the `rmarkdown` R package. It can be knitted to run R code chunks, producing a document with the code results, and converted into another file format such as hypertext markup language (HTML), thanks to the `knitr` R package.

A shiny app serves as an interactive web interface running R code, allowing the user to provide the required input for analyses and to run the RMD scripts. The shiny app consists in a shiny app function `shinyApp(ui, server)` that has 2 arguments: the `ui` object that defines the user interface (UI) and the `server` function processed by the app.

Lastly, functions created specifically for ABSP are stored in a separate file of functions and are called within the RMD script of each analysis.

Hence, these 4 main files constitute the bulk of the ABSP code:

- An **RMD script for the individual analysis**, named *ABSP_individual_analysis.Rmd* that carries out the individual analysis and generates an analysis report in an HTML file. This script is 2,600 lines long and contains 160 chunks of embedded R code.
- An **RMD script for the grouped analysis**, that carries out the grouped analysis and generates an analysis report in an HTML file. This script is 1,600 lines long and contains 38 chunks of embedded R code.
- A **shiny app R script**, it is a 1,500 lines long R script named *app.R*, containing the shiny app function `shinyApp(ui, server)` along with its `ui` and `server` arguments.
- An **R script of functions**, which contains a list of 27 created functions, used in the three previous scripts, over 1,600 lines of code.

2.3.3 APP INTERFACE AND USABILITY

The implemented shiny app provides a user interface, opening in a web browser, aiming to facilitate the input entry and the launching of analyses for users. It allows users unaccustomed to R and RStudio to easily run analyses.

The interface contains 4 tabs:

1. A **main** tab.
2. An **individual analysis** tab.
3. A **grouped analysis** tab.
4. A **multiple analyses** tab.

What is ABSP ?

ABSP is a R based tool to analyze Bisulfite Sequencing PCR (BSP) experiment results, which is why it is the acronym for 'Analysis of Bisulfite Sequencing PCR'. It was developed to help researchers estimate and compare CpG methylation percentages of a DNA region studied using BSP experiments. It provides a complete automated workflow, from trace file sequencing results to data visualization and statistics.

- Complete workflow from raw data to statistics
- Fully automated and user-friendly
- Analysis of Direct-BSP and Cloning-BSP
- Accessibility and flexibility using the R language

Please, cite

Denoulet et al., 2022

Ressources

- The ABSP tool is available for download on github at <https://github.com/ABSP-methylation-tool/ABSP>.
- For detailed instructions, please find the user guide in your ABSP folder.

Contact

absp@univ-lille.fr

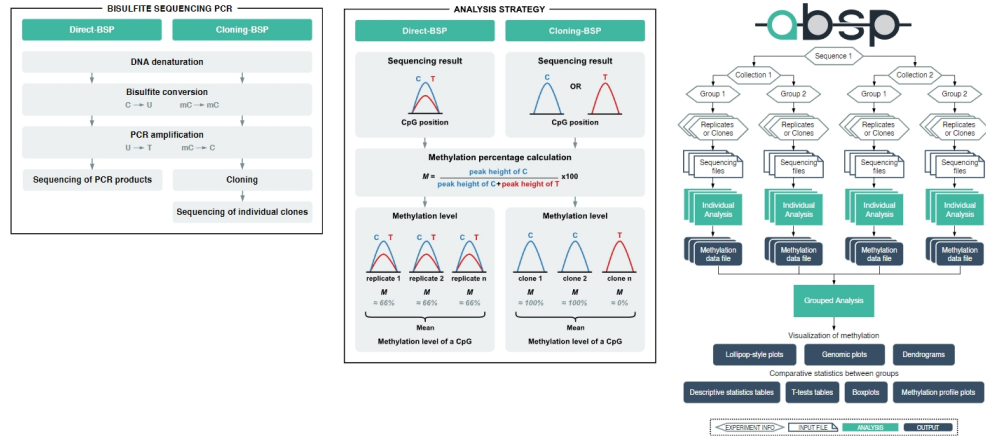


FIGURE 50 Panel 1. ABSP app main tab.

obsp Analysis of Bisulfite Sequencing PCR

Main Individual analysis Grouped analysis Multiple analyses

Sample information

Select existing folder

Select existing sequence folder

Enter collection name

Enter group name

Direct-BSP
 Enter replicate number

Cloning-BSP
 Enter clone number

Reference DNA sequence

Select genome
 Pre-install genome

Select .fasta file of reference DNA sequence
 Upload complete

Sequencing results

Select date of sequencing #1

Select date of sequencing #2

Select sequencing file #1
 Reset

Select sequencing file #2
 Reset

Uploaded file: CDH1 High methylated DNA rep1 Forward.ab1

Uploaded file: CDH1 High methylated DNA rep1 Reverse.ab1

▶ Run individual analysis

Sample information | Reference DNA sequence | Sequencing results

Be aware that these entries must not contain any special characters: / \ : * ? ! " ' ` < > | & % @ # + = { }

Allowed characters (that must be avoided if not necessary): space - _

All entries must be named consistently between the different analysis (e.g. the group names must strictly be identical between samples of the same group).

Select existing folder

Select an existing folder within the ABSP results folder to locate all of the analysis results. Having different folders of results can be used to separate the different analysis by projects, experiments or users.

To create a new folder, select the 'Create new folder' entry and enter the name of the new folder in the text input. Note that the six first letters will appear in the report file name.

Select existing sequence folder

Select an existing folder corresponding to the name of your sequence. This is the name that will be used on tables and plots to refer to the sequence.

To analyze a new sequence and therefore create a new sequence folder, select the 'Create new sequence folder' entry and enter the new sequence name in the text input.

Enter collection name

The collection corresponds to a separation of samples above groups. Samples from different collections can not be compared, even if they belong to the same group. For example, collections can be different cell lines, organs, or patients, in which the same groups are compared but not between the different collections. To compare these types of samples, consider them as groups.

If you do not want to specify any collection, leave empty or enter '0'. Make sure the collection name is strictly identical for all samples of the same collection.

Enter group name

The group corresponds to the condition you want to compare in the grouped analysis. For example, groups can be the 'control' and 'treated' conditions.

The group field entry is required. Make sure the group name is strictly identical for all samples of the same group.

Direct-BSP - Enter replicate number

In the case of direct sequencing of PCR products only (direct-BSP). The replicate number corresponds to the number of repetition for one group, for experiment reproducibility and statistic significance determination. A minimum of 3 replicates is generally recommended.

Cloning-BSP - Enter clone number

In the case of clone sequencing only (cloning-BSP). The clone number corresponds to the identification number of each clone for one group. In general, the sequencing of 10 clones is recommended for results reliability.

ABSP v1.0.0 - Copyright © 2022 CANTHER laboratory, released under the GPL-3 license Close

FIGURE 50 Panel 2. ABSP app individual analysis tab.

Main
Individual analysis
Grouped analysis
Multiple analyses

Experiment information

Select folder

Select sequence

Select genome

Select experiment type
 Direct-BSP Cloning-BSP

Plot parameters

Select position labels for plots
 CpG coordinates CpG numbers None

Collection separation
 Separate plots by collection

Indicate the order of groups for display (all must be selected)

Select the types of sample ordering for plots (multiple choice allowed)

▶ Run grouped analysis

Experiment information [Plot parameters](#)

Select folder
 Select an existing folder within the ABSP results folder to retrieve the previously generated methylation data files after individual analyses, and to locate the newly generated results of the grouped analysis.

Select sequence
 Select an existing folder corresponding to the name of your sequence, to retrieve the previously generated methylation data files after individual analyses, and to locate the newly generated results of the grouped analysis.

Select genome
 Select the reference genome. Only used to display the genomic sequence in the genomic plots.

Select experiment type
 Select the experiment type, either 'Direct-BSP' or 'Cloning-BSP', for proper retrieval of methylation data and analysis specification.

ABSP v1.0.0 - Copyright © 2022 CANTHER laboratory, released under the GPL-3 license
Close

FIGURE 50 Panel 3. ABSP app grouped analysis tab.

obsp Analysis of Bisulfite Sequencing PCR

Main Individual analysis Grouped analysis Multiple analyses

Select existing folder

Example data ▼

Select individual analyses table file

Browse... multiple_individual_analyses_table - CDH1.xlsx Reset

Upload from file

Uploaded file: multiple_individual_analyses_table - CDH1.xlsx

Select grouped analyses table file

Browse... multiple_grouped_analyses_table - CDH1.xlsx Reset

Upload from file

Uploaded file: multiple_grouped_analyses_table - CDH1.xlsx

▶ Run analyses

Principle Diagram of analysis launch

Automated launch of multiple analyses

Several analyses can be launched from this tab. In only one click, using pre-filled tables containing the input entries.

How to proceed ?

1. Fill one or the two input tables below with the sample information and the choice of parameters (you can open the tables files with the buttons below).
2. In the left panel, select an existing folder within the ABS P results folder to locate all of the analyses results.
- To create a new folder, select the 'Create new folder' entry and enter the name of the new folder in the text input. Note that the six first letters will appear in the report file name.
3. In the left panel, select your filled table as input.

Both the sample data table for multiple individual analyses and the parameters table for multiple grouped analyses can be provided at the same time to launch the individual analyses followed by the grouped analyses, or only one of the two tables can be provided and will launch the corresponding analyses, either individual analyses or grouped analyses.

4. Launch the analyses by clicking on the bottom button 'Run analyses'.

All the input analyses will be launch one after the other, do not close the app until the end.

Note: The tables files are located in the 'documents' folder. The input file format must be either 'xlsx' or 'csv'.

Sample data for individual analyses

The table regroupes the individual analyses inputs.

📁 Open multiple individual analysis table file

Parameters for the grouped analyses

The table regroupes the grouped analyses inputs.

📁 Open multiple grouped analyses table file

ABSP v1.0.0 - Copyright © 2022 CANTHER laboratory, released under the GPL-3 license
Close

FIGURE 50 Panel 4. ABSP app multiple analyses tab.

FIGURE 50 ABSP APP INTERFACE. **Panel 1.** Main tab or home page of the app, it provides key links and resources for the users. **Panel 2.** Individual analysis tab, it serves to launch the analysis of a single sample using two sequencing files, one for each direction, to obtain CpG methylation percentages. On the left, input information and files must be provided in the dedicated text and file upload boxes. **Panel 3.** Grouped analysis tab, it serves to launch the analysis of all sample methylation results after their individual analysis to compare CpG methylation percentages between groups. On the left, input information must be selected in the selection boxes. **Panel 4.** Multiple analyses tab, this tab can be used to launch multiple individual and/or grouped analyses at once, using pre-filled input tables that must be provided in the upload file boxes.

All the app tabs are displayed in [Figure 50](#). The ① **main** tab exposes general information and resources. The ② **individual analysis** and ③ **grouped analysis** tabs are dedicated to the launching of the individual analysis and grouped analysis, respectively. They are divided into two parts, one left panel with input boxes and the “Run analysis” button, and one right panel containing subtabs that provide all the necessary guidelines to fill input boxes. The ④ **multiple analyses** tab aims to launch both individual analyses and grouped analyses through the upload of a pre-filled table file containing all the analyses inputs. It is also composed of a left input panel and a right panel with input guidelines. This tab procures a further automated way to run ABSP analyses, more details are given in section 2.3.6 “Multiple analyses” at page 231.

2.3.4 SCRIPT AUTOMATIZATION

To make the analytic process fully automated, input entries from the shiny app are used as parameters for the RMD scripts, which can be entirely run in one go to generate all the analysis results and a full analysis HTML report.

The interface of the shiny app allows the user to specify the desired input parameters and upload data files, thanks to its UI. This input information is then processed by the shiny app server as detailed in [Code listing 5](#), to render the RMD analysis scripts. Indeed, when the “Run analysis” button is pressed (`observeEvent()` function), values of each input entry is processed and stored in a list of parameters (`indiv_params` and `grouped_params`), and passed down in the `rmarkdown::render()` function. This function launches the corresponding RMD script to run and render the HTML (.html) report file.

In the two RMD analysis scripts, these parameters are declared in the YAML header of the document as `params` fields ([Code listing 6](#)). When the rendering of the script is triggered through the shiny app, input entries are assigned to each declared `params` of the RMD script.

Hence, as illustrated in [Code listing 7](#), `params` fields are read in the RMD scripts and stored as objects, allowing the analysis to run with specific parameters.

During the analysis, RMD scripts produce output files such as PNG plots or CSV tables. At the end of the analysis, each RMD script generates an HTML file that gathers all of the analysis

results. A few tabs of the individual analysis and grouped analysis reports are illustrated in **Figure 51**. Examples of reports are available for download at <https://github.com/ABSP-methylation-tool/ABSP> (“examples/reports” folder).

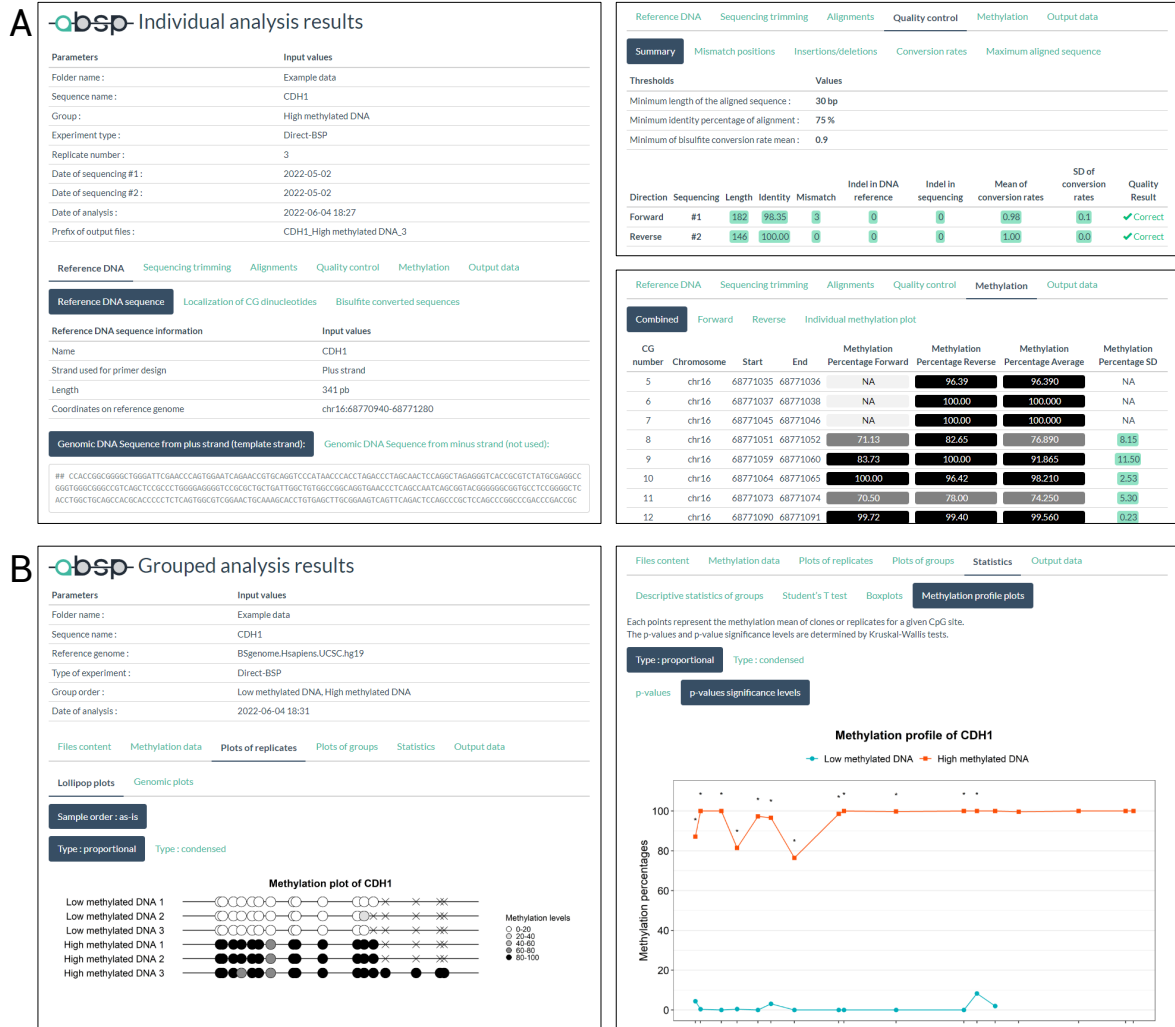


FIGURE 51 REPORTS GENERATED BY THE INDIVIDUAL AND GROUPED ANALYSES. **A.** Example of an HTML report generated by the individual analysis. The upper part contains all the sample information that was inputted when launching the analysis and the bottom part contains the analysis results divided into 6 main tabs: Reference DNA sequence, Sequence trimming, Alignments, Quality control, Methylation, and Output data, themselves subdivided into multiple tabs. **B.** Example of an HTML report generated by the grouped analysis. The upper part contains all the experiment information that was inputted when launching the analysis and the bottom part contains the analysis results divided into 6 main tabs: Files content, Methylation data, Plots of replicates (or Plots of clones for cloning-BSP experiments), Plots of groups, Statistics, and Output data, themselves subdivided into multiple tabs.

2.3.5 ERROR HANDLING

An error occurring during the knitting of an RMD script is stopping the analysis without generating a report file of the analysis. As an example, the first potential error that can cause a knit failure

```

1012 #-----
1013 # Render Rmarkdown report of analysis -----
1014
1015 observeEvent(input$indiv, {
1016   output$indivReport <- renderUI(NULL)
1017
1018   # Get folder name
1019   if(input$folderexisting == "Create new folder") { foldername <- input$foldernew } else { foldername <- input$folderexisting }
1020
1021   # Get sequence folder name
1022   if(input$seqexisting == "Create new sequence folder") { seqname <- input$seqnew } else { seqname <- input$seqexisting }
1023
1024   # Sample name based on input
1025   indiv_name <- paste(substr(foldername, 1, 6), # folder name 6 first characters
1026                     seqname,
1027                     ifelse(input$collection=="0" | input$collection=="", "", input$collection), # collection name if exists
1028                     input$group,
1029                     ifelse(input$replicate=="0" | input$replicate=="", "", input$replicate),
1030                     ifelse(input$clone=="0" | input$clone=="", "", input$clone), sep="-")
1031   indiv_name <- gsub("-", "", indiv_name)
1032   indiv_name <- gsub("$", "", indiv_name)
1033
1034   # Name of report file [ABSP-Indiv]_[folder-seqname-collection-group-replicate-clone]_[YYYYMMDD-HHMM].html
1035   indiv_file <- paste0(paste("ABSP-Indiv", indiv_name, format(Sys.time(), "%Y%m%d-%H%M"), sep = "_"), ".html")
1036
1037   # Path of report file
1038   if (!dir.exists(file.path(getwd(), "reports"))) {suppressWarnings(dir.create(file.path(getwd(), "reports")))}
1039   indiv_path <- file.path(getwd(), "reports", indiv_file)
1040
1041   # Set up parameters to pass to Rmd document
1042   indiv_params <- list(foldername = foldername,
1043                      seqname = seqname,
1044                      collection = input$collection,
1045                      group = input$group,
1046                      replicate = input$replicate,
1047                      clone = input$clone,
1048                      date_s1 = input$date_s1,
1049                      date_s2 = input$date_s2,
1050                      genome = input$genomeI,
1051                      DNA_seq = upload_DNA_seq(),
1052                      ab1_s1 = ab1_s1_file(),
1053                      ab1_s2 = ab1_s2_file())
1054
1055   indiv_RMD <- list.files(path=file.path(getwd(), "scripts"), pattern = "ABSP_individual_analysis.Rmd$", full.names = T)
1056
1057   withProgress(message = "Analysis in progress, please wait!", {
1058     rmarkdown::render(indiv_RMD, encoding = 'UTF-8', output_file = indiv_path, params = indiv_params, envir = new.env(parent = globalenv()),
1059                      on.exit({
1060                        output$indivReport <- renderUI({
1061                          fluidRow(
1062                            align="center",
1063                            br(),
1064                            p("Analysis done!", style="font-size:17px ;"),
1065                            p("The report file ", em(basename(indiv_path)), " has been generated in the ", em("reports"), " directory", style="font-size:17px ;")
1066                          )
1067                        })
1068                      })
1069   })
1070 }
1071 }
1072 }
1073 }
1074 #-----

```

```

1125 #-----
1126 # Render Rmarkdown report of analysis -----
1127
1128 observeEvent(input$grouped, {
1129
1130   # Experiment name based on input
1131   grouped_name <- paste(substr(input$folderselect, 1, 6), # folder name 6 first characters
1132                       input$seqselect,
1133                       ifelse(input$exptype=="Direct-BSP", "directBSP", "cloningBSP"), sep="-")
1134
1135   # Name of report file [ABSP-Indiv]_[folder-seqname-collection-group-replicate-clone]_[YYYYMMDD-HHMM].html
1136   grouped_file <- paste0(paste("ABSP-Grouped", grouped_name, format(Sys.time(), "%Y%m%d-%H%M"), sep = "_"), ".html")
1137
1138   # Path of report file
1139   if (!dir.exists(file.path(getwd(), "reports"))) {suppressWarnings(dir.create(file.path(getwd(), "reports")))}
1140   grouped_path <- file.path(getwd(), "reports", grouped_file)
1141
1142   # Set up parameters to pass to Rmd document
1143   grouped_params <- list(foldername = input$folderselect,
1144                        seqname = input$seqselect,
1145                        genome = input$genomeG,
1146                        cloning = input$exptype,
1147                        coll_sep = input$coll_sep,
1148                        pos_labels = input$pos_labels,
1149                        group_order = paste(input$group_order, collapse = ","),
1150                        sample_order = input$sample_order)
1151
1152   grouped_RMD <- list.files(path=file.path(getwd(), "scripts"), pattern = "ABSP_grouped_analysis.Rmd$", full.names = T)
1153
1154   withProgress(message = "Analysis in progress, please wait!", {
1155     rmarkdown::render(grouped_RMD, encoding = 'UTF-8', output_file = grouped_path, params = grouped_params, envir = new.env(parent = globalenv()),
1156                      on.exit({
1157                        output$groupedReport <- renderUI({
1158                          fluidRow(
1159                            align="center",
1160                            br(),
1161                            p("Analysis done!", style="font-size:17px ;"),
1162                            p("The report file ", em(basename(grouped_path)), " has been generated in the ", em("reports"), " directory", style="font-size:17px ;")
1163                          )
1164                        })
1165                      })
1166   })
1167 }
1168 }
1169 }
1170 }
1171 #-----

```

CODE LISTING 5 *Caption in following pages.*

```

1322 # -----
1323 # Run all individual analysis from the sample data table -----
1324
1325 list_Ireports <- c()
1326
1327 for (i in 1:nrow(exp_table)) {
1328
1329
1330 # Sample name based on idata table
1331 indiv_name <- paste(substr(foldermain, 1, 6), # folder name 6 first characters
1332                   exp_table[i,1],
1333                   ifelse(exp_table[i,2]=="0" | is.na(exp_table[i,2]), "", exp_table[i,2]), # collection name if ex
1334                   exp_table[i,3],
1335                   ifelse(exp_table[i,4]=="0" | is.na(exp_table[i,4]), "", exp_table[i,4]),
1336                   ifelse(exp_table[i,5]=="0" | is.na(exp_table[i,5]), "", exp_table[i,5]), sep="-")
1337 indiv_name <- gsub("--", "-", indiv_name)
1338 indiv_name <- gsub("$", "", indiv_name)
1339
1340 # Name of report file [ABSP-Indiv]_[folder-seqname-collection-group-clone]_[YYYYMMDD-HHMM].html
1341 indiv_file <- paste0(paste("ABSP-Indiv", indiv_name, format(Sys.time(), "%Y%m%d-%H%M"), sep = "-"), ".html")
1342
1343 # Path of report file
1344 indiv_path <- file.path(output_rep, indiv_file)
1345
1346 # Set up parameters to pass to Rmd document
1347 indiv_params <- list(foldername = foldermain,
1348                   seqname = exp_table[i,1],
1349                   collection = ifelse(is.na(exp_table[i,2]), exp_table[i,2], ""),
1350                   group = exp_table[i,3],
1351                   replicate = as.integer(ifelse(is.na(exp_table[i,4]), exp_table[i,4], "")),
1352                   clone = as.integer(ifelse(is.na(exp_table[i,5]), exp_table[i,5], "")),
1353                   genome = exp_table[i,6],
1354                   DNA_seq = exp_table[i,7],
1355                   date_s1 = exp_table[i,8],
1356                   ab1_s1 = exp_table[i,9],
1357                   date_s2 = exp_table[i,10],
1358                   ab1_s2 = exp_table[i,11])
1359
1360 indiv_RMD <- list.files(path=file.path(getwd(), "scripts"), pattern = "ABSP_individual_analysis.Rmd$", full.names = T)
1361
1362 withProgress(message = "Analysis in progress, please wait!", {
1363   rmarkdown::render(indiv_RMD, encoding = 'UTF-8', output_file = indiv_path, params = indiv_params)
1364   on.exit({
1365     output$multipleIReport <- renderUI({
1366       fluidRow(
1367         align="center",
1368         br(),
1369         p(basename(indiv_path), " has been generated in the ", em("reports"), " directory", style="font-size:15px ;")
1370       )
1371     })
1372   })
1373 }

```

```

1404 # -----
1405 # Run all grouped analysis of previously analysed experiments -----
1406
1407 list_Greports <- c()
1408
1409 for (i in 1:nrow(grouped_table)) {
1410
1411 # Experiment name based on input
1412 grouped_name <- paste(substr(foldermain, 1, 6), # folder name 6 first characters
1413                   grouped_table[i,1],
1414                   ifelse(grouped_table[i,3]=="Direct-BSP", "directBSP", "cloningBSP"), sep="-")
1415
1416 # Name of report file [ABSP-Indiv]_[folder-seqname-collection-group-clone]_[YYYYMMDD-HHMM].html
1417 grouped_file <- paste0(paste("ABSP-Grouped", grouped_name, format(Sys.time(), "%Y%m%d-%H%M"), sep = "-"), ".html")
1418
1419 # Path of report file
1420 grouped_path <- file.path(getwd(), "reports", grouped_file)
1421
1422 # Split string by commas to get list of sample orders
1423 sample_order <- strsplit(grouped_table[i,7], ",")[1]
1424 sample_order <- gsub("^ ", "", sample_order) # remove first spaces
1425 sample_order <- gsub(" $", "", sample_order) # remove last spaces
1426
1427 # Set up parameters to pass to Rmd document
1428 grouped_params <- list(foldername = foldermain,
1429                   seqname = grouped_table[i,1],
1430                   genome = grouped_table[i,2],
1431                   cloning = grouped_table[i,3],
1432                   pos_labels = grouped_table[i,4],
1433                   coll_sep = grouped_table[i,5],
1434                   group_order = grouped_table[i,6],
1435                   sample_order = sample_order)
1436
1437 grouped_RMD <- list.files(path=file.path(getwd(), "scripts"), pattern = "ABSP_grouped_analysis.Rmd$", full.names = T)
1438
1439 withProgress(message = "Analysis in progress, please wait!", {
1440   rmarkdown::render(grouped_RMD, encoding = 'UTF-8', output_file = grouped_path, params = grouped_params,
1441                     envir = new.env(parent = globalenv()))
1442   on.exit({
1443     output$multipleIReport <- renderUI(NULL)
1444     output$multipleGReport <- renderUI({
1445       fluidRow(
1446         align="center",
1447         br(),
1448         p(basename(grouped_path), " has been generated in the ", em("reports"), " directory", style="font-size:15px ;")
1449       )
1450     })
1451   })
1452 }

```

CODE LISTING 5 *Caption next page.*

CODE LISTING 5 APP SCRIPT: RENDERING INDIVIDUAL, GROUPED, AND MULTIPLE ANALYSES. **A.** Rendering the individual analysis from inputs of the individual analysis tab. **B.** Rendering the grouped analysis from inputs of the grouped analysis tab. **C.** Rendering all individual analyses from the input table of the multiple analyses tab. **D.** Rendering all grouped analyses from the input table of the multiple analyses tab.

is the incorrect input of genomic coordinates, not matching the length of the provided reference DNA sequence (Code listing 8). Indeed, the `seq_start` and `seq_end` variables are used in the code and their difference must correspond to the reference sequence length (as the first nucleotide is always numbered as 1 and not 0, 1 must be added to the difference in positions to obtain the length). To handle this error and allow the analysis to properly stop with the generation of a truncated report, the matching between coordinates and sequence length is checked before the error occurrence. If the matching is incorrect, the `TRUE` value is stored in the `genocoord` object, and induces the next chunks to be run — as they are conditioned by the `eval=genocoord` value — which stops the analysis and generating the truncated HTML report. A list of errors is available in the troubleshooting section of the ABSP user guide in Appendix 3 “ABSP user guide” at page 338.

2.3.6 MULTIPLE ANALYSES

To push the automatization further, especially when dealing with numerous samples, an additional tab, named “multiple analyses”, has been added to shorten the process of launching several analyses at once while optimizing the entry of inputs (Figure 50 panel 4 and Figure 52). Two spreadsheet documents can be filled with input information required for both the individual analyses and grouped analyses. Input entries required in the individual analysis and grouped analysis tabs are therefore transposed as columns of variables to fill within the spreadsheet, with each row corresponding to each sample in the individual analyses document, or to a grouped analysis of a unique sequence in the grouped analyses document. For files, paths to input files have to be entered within the corresponding columns. Guidelines and required formats are described within the spreadsheet documents to help users.

Input tables are then uploaded into the multiple analyses tab to launch individual analyses and grouped analyses, either both at once or separately. The app script imports the spreadsheet input tables and uses their data as parameters to launch the successive analyses.

In conclusion, the ABSP tool constitutes a new analytic tool to help researchers interpret results from BSP data. By providing a ready-to-use solution to analyze these data, ABSP improves access to the study of DNA methylation at target regions. Indeed, as the BSP technique is affordable and accessible, coupled with a result analysis by ABSP, this workflow allows to rapidly get an evaluation of methylation levels at a specific DNA region.


```

A
8  params:
9  foldername:
10     label: "Enter folder name"
11     value: folder-name
12     input: text
13  seqname:
14     label: "Enter sequence name"
15     value: sequence-name
16     input: text
17  collection:
18     label: "Enter collection name"
19     value: collection
20     input: text
21  group:
22     label: "Enter group name"
23     value: group1
24     input: text
25  replicate:
26     label: "Enter replicate number (direct-BSP)"
27     value: 1
28     input: numeric
29  clone:
30     label: "Enter clone number (cloning-BSP)"
31     value: 0
32     input: numeric
33  DNA_seq:
34     label: "Select .fasta file of reference DNA sequence from plus strand with header containing coordinates and strand
      (plus or minus) used for primer design"
35     value: "DNA_seq.fasta"
36     input: file
37  genome:
38     label: "Select genome"
39     value: BSgenome.Hsapiens.UCSC.hg19
40     input: select
41     choices: !r installed.genomes()
42     selected: BSgenome.Hsapiens.UCSC.hg19
43  date_s1:
44     label: "Select date of sequencing #1"
45     value: !r Sys.Date()
46     input: date
47  date_s2:
48     label: "Select date of sequencing #2"
49     value: !r Sys.Date()
50     input: date
51  ab1_s1:
52     label: "Select .ab1 file of sequencing result from one direction #1"
53     value: "seq_s1.ab1"
54     input: file
55  ab1_s2:
56     label: "Select .ab1 file of sequencing result from the other direction #2"
57     value: "seq_s2.ab1"
58     input: file

B
8  params:
9  foldername:
10     label: "Enter folder name from the results folder"
11     input: text
12     value: folder-name
13  seqname:
14     label: "Enter sequence name"
15     input: text
16     value: sequence-name
17  genome:
18     label: "Select genome"
19     input: select
20     choices: !r installed.genomes()
21     selected: BSgenome.Hsapiens.UCSC.hg19
22     value: BSgenome.Hsapiens.UCSC.hg19
23  cloning:
24     label: "Select experiment type"
25     input: radio
26     choices: [Direct-BSP, Cloning-BSP]
27     value: Direct BSP
28  coll_sep:
29     label: "Separate plots by collection"
30     input: checkbox
31     value: FALSE
32  pos_labels:
33     label: "Type of position labels for plots"
34     input: select
35     choices: [CpG coordinates, CpG numbers, None]
36     selected: CpG coordinates
37     value: CpG coordinates
38  group_order:
39     label: "Indicate the list of groups in the correct order (comma separated)"
40     input: text
41     value: group1, group2, group3
42  sample_order:
43     label: "Select the type(s) of sample ordering (multiple choice allowed)"
44     input: select
45     multiple: TRUE
46     choices: [As it is, By groups, By methylation levels, By clusters]
47     selected: As it is
48     value: [As it is, By groups, By methylation levels, By clusters]

```

CODE LISTING 6 PARAMETERS IN YAML HEADERS OF INDIVIDUAL AND GROUPED ANALYSES SCRIPTS. **A.** Parameters for the individual analysis. **B.** Parameters for the grouped analysis.

A

```

103 ~ `r Read parameters from YAML header, include=F}
104 ## SCRIPT TO USE PARAMS FOR KNITING : DO NOT RUN IF TESTING THE SCRIPT WITHOUT KNITTING
105
106 # Parameters variables
107 folder_name <- params$foldername
108 seq_name <- params$seqname
109 seq_coll <- params$collection ; seq_coll <- ifelse(seq_coll==0 | is.na(seq_coll),"",seq_coll)
110 seq_group <- params$group
111 if (params$replicate==0 | is.na(params$replicate)) {seq_rep <- ""} else {seq_rep <- params$replicate}
112 if (params$clone==0 | is.na(params$clone)) {seq_clone <- ""} else {seq_clone <- paste0("clone",str_pad(params$clone, 2, "0",
113   side="left"))}
114 if (params$clone==0 | is.na(params$clone)) {clone_nb <- NA} else {clone_nb <- params$clone}
115
116 # Date
117 if(length(params$date_s1)==0) {seq_date_s1 <- ""}
118 } else if (is.null(params$date_s1) | is.na(params$date_s1)) { seq_date_s1 <- ""}
119 } else {seq_date_s1 <- params$date_s1}
120
121 if(length(params$date_s2)==0) {seq_date_s2 <- ""}
122 } else if (is.null(params$date_s2) | is.na(params$date_s2)) { seq_date_s2 <- ""}
123 } else {seq_date_s2 <- params$date_s2}
124
125 # Date format
126 seq_date_s1 <- as.character(seq_date_s1)
127 seq_date_s2 <- as.character(seq_date_s2)
128
129 selected_genome <- params$genome
130
131 # Files
132 if(length(params$ab1_s1)==0) { useSeqs1 <- F}
133 } else if (params$ab1_s1=="seq-s1.ab1" | is.na(params$ab1_s1) | is.null(params$ab1_s1) | params$ab1_s1=="") { useSeqs1 <- F}
134 } else { useSeqs1 <- T}
135
136 if(length(params$ab1_s2)==0) { useSeqs2 <- F}
137 } else if (params$ab1_s2=="seq-s2.ab1" | is.na(params$ab1_s2) | is.null(params$ab1_s2) | params$ab1_s2=="") { useSeqs2 <- F}
138 } else { useSeqs2 <- T}
139
140 DNA_seq_path <- params$DNA_seq
141 sangerseq_s1_path <- params$ab1_s1
142 sangerseq_s2_path <- params$ab1_s2
143
144 ~ `r

```

B

```

134 ~ `r Read parameters from YAML header,include=FALSE}
135 ## SCRIPT TO USE PARAMS FOR KNITING : DO NOT RUN IF TESTING THE SCRIPT WITHOUT KNITTING
136
137 # Parameters variables
138
139 folder_name <- params$foldername
140 seq_name <- params$seqname
141 selected_genome <- params$genome
142
143 # Type of experiment
144 if (params$cloning == "Cloning-BSP") {
145   cloning_exp <- TRUE
146 } else if (params$cloning == "Direct-BSP"){
147   cloning_exp <- FALSE}
148
149 # Parameter to separate collections on different graphs if TRUE
150 coll_sep <- params$coll_sep
151
152 # Change annotation of position labels format
153 pos_labels <- params$pos_labels %>%
154   replace(params$pos_labels == "CpG coordinates","coordinates") %>%
155   replace(params$pos_labels == "CpG numbers","numbers") %>%
156   replace(params$pos_labels == "None","none")
157
158 # Split string by commas to get list of groups
159 GroupOrder <- strsplit(params$group_order,",")
160 GroupOrder <- GroupOrder[[1]]
161 # Removes start/end spaces in group order string
162 GroupOrder <- gsub("^ ","",GroupOrder)
163 GroupOrder <- gsub("$"," ",GroupOrder)
164
165 # Change annotation of sample orders
166 list_sampleOrder <- params$sample_order %>%
167   replace(params$sample_order == "As it is","as-is") %>%
168   replace(params$sample_order == "By groups","by-group") %>%
169   replace(params$sample_order == "By methylation levels","by-meth") %>%
170   replace(params$sample_order == "By clusters","clust")
171 ~ `r

```

CODE LISTING 7 READING PARAMETERS IN INDIVIDUAL AND GROUPED ANALYSES SCRIPTS. A. Reading parameters in the individual analysis. **B.** Reading parameters in the grouped analysis.

```

448- {r check if incorrect genomic coordinates}
449- # Verify correct genomic coordinates
450- if ( (seq_end - seq_start) - length(DNA_seq) != -1 ) {genocoord <-T} else {genocoord <-F}
451-
452-
453- {r output if incorrect genomic coordinates, results='asis', eval=genocoord}
454- # Output tab and Directories tab
455- cat("## Output data {.tabset .tabset-pills}","## Directories", sep=" \n")
456-
457-
458- {r diagram if incorrect genomic coordinates, eval=genocoord}
459- # Diagram of directories
460- diagram_indiv
461-
462-
463- {r stop knit if incorrect genomic coordinates, results='asis', eval=genocoord}
464- # Files tab with list of output files
465- cat("## Files", " \n")
466- list_output_indiv(fileprefix = fileprefix)
467- # Abort analysis and error message
468- cat("## Analysis aborted", " \n")
469- p("Error: Length of reference DNA sequence does not match with the provided genomic coordinates. Please verify concordance
470- between the reference DNA sequence and genomic coordinates.", style="color:#bf3232 ; font-weight: bold ;")
471- shiny::setProgress(1)
472- knit_exit()

```

CODE LISTING 8 HANDLING INCORRECT GENOMIC COORDINATES ERROR IN THE INDIVIDUAL ANALYSIS SCRIPT.

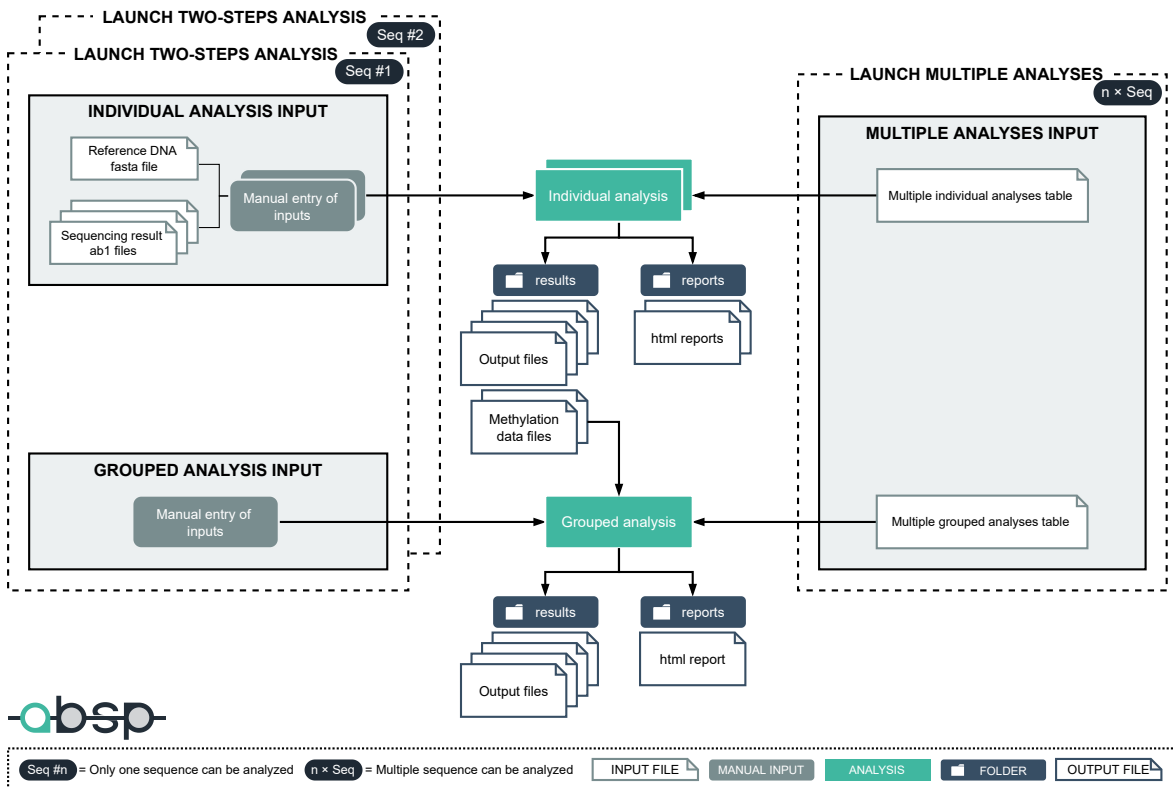


FIGURE 52 MULTIPLE WAYS TO LAUNCH THE ABSP ANALYSES.

KEY POINTS

- ➔ ABSP provides a complete analytic process, from raw BSP data to comparative analysis. It controls the quality of sequencing data, calculates CpG methylation levels from each sample, generates publication-ready graphical representations for methylation visualization, and compares groups of samples using statistical tests to determine significant methylation differences, and helps researchers answer the experimental hypothesis.
- ➔ ABSP is fully automated and user-friendly from launch to result interpretation. A built-in shiny app provides a web interface to guide the user to launch analyses. Output files and analysis HTML reports are automatically saved in the user's folders.
- ➔ ABSP can analyze both BSP approaches: direct-BSP and cloning-BSP. The workflow developed to analyze direct-BSP data have been transposed to also analyze cloning-BSP, providing continuity throughout the analytic process.
- ➔ ABSP is accessible and flexible. By being coded in the R programming language, a free and widely used software, the source code of ABSP is available and can be easily modified by R accustomed users, to adapt the software to experimental needs.
- ➔ ABSP is available for download, along with its user manual and associated example data at <https://github.com/ABSP-methylation-tool/ABSP>.

^ [Back to Table of Contents](#)

DISCUSSION AND PERSPECTIVES

DISCUSSION AND PERSPECTIVES

☰ CHAPTER CONTENTS

1	IDENTIFICATION OF DNA METHYLATION CHANGES INVOLVED IN RADIO-INDUCED DEDIFFERENTIATION	239
1.1	GLOBAL EXPRESSION ANALYSIS OF EPIGENETIC ENZYMES	239
1.2	CONTRIBUTION OF DNMTs IN THE RADIO-INDUCED DEDIFFERENTIATION	240
1.3	ANALYSIS METHOD OF RRBS DATA TO IDENTIFY DIFFERENTIALLY METHYLATED REGIONS	241
1.4	DIFFERENTIALLY METHYLATED REGIONS AND ASSOCIATED GENES	244
2	THE ABSP TOOL: APPLICATION, LIMITATIONS, AND IMPROVEMENTS	247

[^ Back to Table of Contents](#)

1 IDENTIFICATION OF DNA METHYLATION CHANGES INVOLVED IN RADIO-INDUCED DEDIFFERENTIATION

The hypothesis is that epigenetic modifications occur following radiation exposure, contributing to changes in gene transcription programs in response to irradiation, and allowing the reacquisition of stem cell features in a small subset of cancer cells.

1.1 GLOBAL EXPRESSION ANALYSIS OF EPIGENETIC ENZYMES

As a first step, I decided to get insights into the involvement of epigenetic enzymes in radio-induced dedifferentiation. Indeed, [Antwih *et al.* \(2013\)](#) found out that DNMT1 protein levels were gradually decreasing within 72 hours after 6 Gy irradiation, in MDA-MB-231 cells, another breast cancer cell line. A decrease of the DNMT enzymes expression could result in global DNA hypomethylation, which is known to be a key feature of ESC epigenetic reprogramming during early embryogenesis, and thus, could be implicated in cancer cell phenotypic switches in the same way. Additionally, as epigenetic marks are linked and complementary, histones modifying enzymes, such as KDMs or EZH2 could also be implicated in resetting the epigenetic landscape of genes related to stemness.

The expression of several epigenetic associated enzymes, DNMT1, DNMT3A, DNMT3B, TET1, TET2, TET3, KDM6A, KDM6B, and EZH2, was assessed following a radiotherapy treatment both within 5 days post-irradiation in SUM159PT cells and within the first 12 hours post-irradiation in the SUM159PT non-CSC (ALDH⁻) population. This expression analysis was not conclusive as no significant difference was found between irradiated and unirradiated cells. Yet, these findings do not rule out the involvement of these enzymes in the radio-induced dedifferentiation process.

The expression analysis carried out to study the contribution of epigenetic enzymes in the non-CSC-to-CSC phenotypic switch was not the optimal approach, but the first one to get a glimpse of expression level variations occurring during this process. Indeed, because less than 5% of CSCs are generated after irradiation, even if substantial changes in expression occur in these switching cells, expression levels are still diluted in the total population, making any expression changes nearly undetectable. This highlights the main technical issues encountered when studying rare cellular events: it is not possible to isolate cells capable to dedifferentiate from the rest of the cell population, as we cannot predict which cell will acquire stem cell properties.

To evaluate expression modulations related to these rare phenotypic conversion events, two main approaches can be used. One approach is to compare the prior and subsequent states, which is experimentally convenient and accessible but limited by material quantities — with 2.5 to 5% of induced CSCs, a large number of initial cells and several hours of sorting are required to obtain limited amounts of cells — however, it does not allow for the study of dynamic changes during the conversion. The other approach is the single-cell live tracking of these events with fluorescence reporters to dynamically observe expression modulations throughout the phenotypic conversion, as well as associated modifications leading to it. However, this approach has numerous limitations in terms of accessibility and development, and is not always the most suited one, especially for exploratory studies.

Finally, these epigenetic-associated enzymes may be involved in phenotypic switches not by expression level changes but by being differentially recruited at specific genomic regions with the help of co-activators and co-repressors. To evaluate the latter, two approaches can be considered: the inhibition of these enzymes to assess their contribution to the radio-induced dedifferentiation, and the identification of regions undergoing changes in their epigenetic landscape.

This project then focused on DNA methyltransferases (DNMTs) and DNA methylation marks at CpG sites and their involvement in the phenotypic conversion from the non-CSC state to the CSC state.

1.2 CONTRIBUTION OF DNMTs IN THE RADIO-INDUCED DEDIFFERENTIATION

First, to know if the DNMTs activity is required for cells to convert into CSCs by reacquiring stem cell properties, an inhibition approach was carried out. The non-CSC population of SUM159PT cells was transfected with siRNA targeting DNMT1 or DNMT3B and the dedifferentiation was induced by irradiation. The proportion of new CSCs was measured by both a CSC marker assay using flow cytometry (ALDH activity) and a functional sphere-forming capacity (SFC) test 5 days post-irradiation.

A limit of this experiment is the siRNA limited time of efficiency. Indeed, siRNAs are efficient for about 72 hours following transfection, yet the measurement of dedifferentiated CSCs is optimal after 5 days following irradiation which is also 5 days following the siRNA transfection. Yet, the most

important feature in the experimental strategy is the inhibition of DNMTs at the time of irradiation to directly impact the response to radiations and thus it requires to transfect cells before irradiation. Hence, whenever the dedifferentiation event occurs, if its induction is due to the radiation exposure it was necessarily impacted by the DNMT inhibition.

The radio-induced increase of CSCs was abolished in cells transfected with siDNMTs, but it was also abolished in cells transfected with a control siRNA that has no target, indicating that the transfection itself inhibits the generation of new CSCs from non-CSCs. More importantly, the viability of siRNA transfected cells was highly impacted, especially as it is included in a series of stressful steps. Indeed, the combination of FACS sorting followed by irradiation is already affecting the cell viability, but the addition of a siRNA transfection is even more deleterious for cells. For the siRNA transfected cells, their cell viability can generally be estimated from 5 to 20%, which is quite lower than the viability of untransfected cells, estimated at around 40 to 60%. The huge amount of cell death directly affects the cell density and the cell environment, and thus, their phenotypic response is greatly impacted as well and could explain the reduced dedifferentiation event rate. In addition, several other transfection methods (INTERFERin, Lipofectamine, and nanoparticles) were tested but did not improve the cell viability.

Hence, this experimental design did not allow the observation of unbiased phenotypic conversion events in response to radiation exposure due to the deleterious effect of the siRNA transfection. No conclusion can be drawn on the involvement of DNMT enzymes in the radio-induced dedifferentiation into CSC.

To pursue in this direction and resolve this limitation, the stable transfection of cells with short hairpin RNA (shRNA) for example can replace the siRNA transfection step and avoid its associated deleterious effects during the dedifferentiation experiment. As this experimental design is more tedious and time-consuming than the siRNA approach, it was not prioritized in this project.

1.3 ANALYSIS METHOD OF RRBS DATA TO IDENTIFY DIFFERENTIALLY METHYLATED REGIONS

The deciphering of the DNA methylation contribution in the stem-like state reacquisition can be done through the comparison of CpG sites methylation levels between the non-CSC state and the induced CSC state. To picture methylation changes between these two states, the RRBS method was chosen, as it is a good compromise between a reduced cost and a genome-scale coverage of CpG sites, compared to the WGBS method. Indeed, the RRBS technique allows the evaluation of CpG methylation levels on a reduced representation of the genome, enriched in CpG sites (around 10 % of the total CpG sites, and 85% of CGI) thanks to enzymatic digestion (*MspI*, C↓CGG, methylation independent), before bisulfite sequencing (Cokus *et al.*, 2008; Meissner *et al.*, 2005; Smith *et al.*, 2009). The RRBS was therefore carried out on four different subpopulations of SUM159PT cells: CSC, non-CSC, induced CSC by irradiation (iCSC), and irradiated non-CSC (inon-CSC). Once RRBS data are

generated, the aim is to identify methylation differences between these subpopulations, which could reveal regions undergoing methylation changes during radio-induced dedifferentiation. Methylation modifications at these regions could be linked to transcriptomic changes participating in this phenotype conversion. Hence, the identification of methylation changes could lead to the deciphering of genes and mechanisms implicated in the regeneration of new CSCs and could therefore serve as potential therapeutic targets to prevent the CSC enrichment following radiotherapy, thus reducing tumor resistance as well as the risk of recurrence.

RRBS data consists in reading counts of C and T bases at each covered CpG site. Based on the C and T read counts, the methylation level of a single CpG can be calculated as well as the overall region methylation level by grouping a set of CpG sites. Hence, to identify differentially methylated regions (DMRs), the genome has to be subdivided into regions to compute their methylation levels and compare them between subpopulations. I chose two different types of genome subdivisions: a subdivision without *a priori* annotation and a subdivision with *a priori* annotation. For the first type, the genome was divided into sliding tiles (or windows) of both 1,000 bp and 200 bp, which is a basic slicing of the genome without *a priori*. This tiling genome subdivision is the most common approach used in studies containing RRBS/WGBS analyses, and sizes generally used are comprised between 2000 bp and 1,000 bp (Kundu *et al.*, 2021; Song *et al.*, 2013a; Wang *et al.*, 2013b). Additionally, the genome was also subdivided into methylation segments — groups of 10 to 100 adjacent CpG sites with similar methylation levels — thanks to a segmentation algorithm (`methyokit::methSeg()` function based on the `fastseg::fastseg()` function), using methylation data from every sample (Akalın *et al.*, 2012). Then, the second type of regions with *a priori* annotation, two interesting genomic annotations were selected: CpG islands (CGIs) on one side and regions surrounding gene transcription start sites (TSSs) on the other side, located 500 bp upstream and downstream of the TSS (1,001 bp long in total), corresponding to the promoter region and the beginning of the gene. It is pertinent to include the beginning of the gene and not only restrict the genome subdivision to the upstream promoter regions. Indeed, methylation marks on the first exon can also affect the gene transcription, and it is frequent to observe CGIs covering promoters and overlapping the beginning of the gene (Brenet *et al.*, 2011; Jones, 2012; Li *et al.*, 2017b; Vavouri and Lehner, 2012). In addition, Krinner *et al.* (2014) found out that CpG-rich domains located ~700 bp downstream of TSSs are correlated with high transcription levels.

Other annotations could be used as genome subdivisions, such as enhancers (distal cis-regulatory elements) and CGI shores (regions flanking CGIs, commonly up to 2,000 bp in either direction). Indeed, studies have demonstrated that methylation levels of enhancers can be strongly correlated with gene expression (Angeloni and Bogdanovic, 2019; Aran and Hellman, 2013; Hon *et al.*, 2013). Besides, CGI shores can also undergo methylation alterations correlated with gene expression (Doi *et al.*, 2009; Irizarry *et al.*, 2009). For instance, the Caveolin-1 expression in luminal breast cancer is downregulated by shores hypermethylation of the CGI located at the beginning of

the gene, while in basal-like breast cancer, the shores hypomethylation of this CGI is associated with gene expression upregulation (Rao *et al.*, 2013). But, even if these annotations were not specifically investigated during the RRBS data analysis, the tiling subdivision of the genome still allows for a thorough screening of methylation differences, ensuring that enhancers and CGI shores are covered.

Furthermore, the maximum region length was set to 1,000 bp, as wider regions would result in the dilution of methylation differences, impairing DMRs identification. Additionally, methylation segmentation is based on the methylation of adjacent CpG sites, independently of the distance between them. Hence, the number of CpG sites per segment, set from 10 to 100 sites, allows for the investigation of short to very large regions but is relevant in terms of consecutive CpG sites. Altogether, the 5 sets of genomic regions examined are complementary in the process of identifying methylation differences along the genome.

The RRBS analysis was carried out using the R programming language and the `methylKit` R package that offers a great parameter flexibility (e.g. coverage, genomic regions, types of statistical tests) (Akalin *et al.*, 2012). Other tools, either software or R packages, for the analysis of RRBS data are available, such as *BiSeq* (Hebestreit *et al.*, 2013), *DMRfinder* (Gaspar and Hart, 2017), *DSS* (Feng *et al.*, 2014), *HMM-DM* (Yu and Sun, 2016), *metilene* (Jühling *et al.*, 2016), *MethylSig* (Park *et al.*, 2014), and *RRBS-Analyzer* (Wang *et al.*, 2013b), but the `methylKit` package seems to be one of the most used, accessible, and flexible (Li *et al.*, 2018b; Liu *et al.*, 2020d; Wreczycka *et al.*, 2017).

In terms of differential methylation computation, although the `methylKit` package was used as a base pipeline for the analysis, the methylation calculation method applied was not the one implemented by `methylKit` (`methylkit::calculateDiffMeth()` function), providing either a logistic regression test or a Fisher's exact test. Instead, the differential methylation calculation method applied comes from the `DSS` package (`methylkit::calculateDiffMethDSS()` function, integrated in `methylKit`) (Feng *et al.*, 2022; Feng *et al.*, 2014; Feng and Wu, 2019). The three methods were tested on RRBS data and the last one has been selected as it includes biological replicates in the calculation method. The `DSS` method assumes that read count data follows a beta-binomial distribution: a beta distribution among replicates, parameterized with dispersion for biological variations among replicates, and a binomial distribution of methylation proportions based on read counts.

Additionally, on defined regions, differential methylation values are computed based on read counts — number of C (methylated) reads over the number of C (methylated) + T (unmethylated) reads from every CpG site on the region — and not based on methylation levels, as read counts include the sequencing depth information (Feng *et al.*, 2014).

To conclude, the sequencing data obtained from RRBS can be analyzed using different methods, tools, and parameters, each with its benefits and limitations. Here, the choice to use the `methylKit` R package, combined with the `DSS` differential methylation calculation method, and the choice of the different region sets coming from different genome subdivisions, appeared to be the

most pertinent and accessible ones, and even if they can be further improved, they allowed the identification of DMRs by the pairwise comparisons of subpopulations.

1.4 DIFFERENTIALLY METHYLATED REGIONS AND ASSOCIATED GENES

The RRBS analysis aims to identify regions undergoing methylation changes between the non-CSC state and radio-induced CSC state. To do so, differentially methylated regions (DMRs) from three pairwise comparisons, non-CSC vs CSC, CSC vs iCSC, and non-CSC vs iCSC, are generated and then sorted to eliminate the irrelevant ones. The pool of CSC-specific DMRs (from the non-CSC vs CSC comparison) is sorted based on the overlapping with the two other pools of DMRs. Hence, only regions presenting a differential methylation between non-CSC vs CSC states (> 25%) but not between the two CSC states (< 25%) and between the two non-CSC states (< 25%) are kept. It means that these regions present a CSC-specific methylation compared to non-CSCs, which is restored in iCSCs but not in non-CSCs after irradiation. This sorting resulted in the identification of 2,596 regions that are likely to undergo methylation changes during the radio-induced dedifferentiation.

From this list of regions, a filtering is applied based on the differential methylation values from pairwise subpopulations comparisons, in addition to a minimum threshold of 3 CpG sites covered in every sample for robustness. Indeed, regions that match the hypothetical methylation profile of regions undergoing methylation changes specific to the radio-induced dedifferentiation could be implicated in its regulation. This ideal methylation profile is defined by a differential methylation between CSC vs non-CSC states (> 25%) and non-CSC vs iCSC states (> 25%), but not between CSC vs iCSC states (< 10%) and non-CSC vs non-CSC states (< 10%). Hence, these cut-offs applied to the list of 2,596 regions resulted in the filtering of 35 regions that are even more likely to undergo methylation changes during the radio-induced dedifferentiation and may be involved in the regulation of this process.

These 35 differentially methylated regions are associated with 23 unique genes. Among them, the ① ryanodine receptor 2 (RYR2), ② Ras and Rab interactor 2 (RIN2), and ③ forkhead box P1 (FOXP1) genes can be cited as examples. ① The RYR2 gene is located at chr1:237,205,505-237,997,288 coordinates and the 1,000 bp hypomethylated identified DMR (T1000_0078) is located in an intron at chr1:237,235,001-237,236,000 coordinates. RYR2 is a major component of the Ca_2^+ pathway, regulating the Ca_2^+ release from the sarcoplasmic reticulum into the cytoplasm (Ding *et al.*, 2017). In breast cancer, the RYR2 expression was upregulated following EGF-induced EMT in MDA-MB-468 cells, suggesting that its expression is involved in metastasis formation (Davis *et al.*, 2013). ② The RIN2 gene is located at chr20:19,867,165-19,983,101 coordinates and the 200 bp hypomethylated identified DMR (T200_1089) is located in an intron at chr20:19,908,801-19,909,000 coordinates. RIN2 is a key R-Ras mediator that physically and functionally couples R-Ras and Rab5 GTPases, participating in the endothelial cell adhesion to the extracellular matrix, as well as their

migration and vascular morphogenesis (Sandri *et al.*, 2012). ③ The FOXP1 gene is located at chr3:71,003,844-71,633,140 coordinates and is associated with several small hypomethylated DMRs from methylation segments (SEGR_0019, SEGR_0020, and SEGR_0021), all located within a 1,000 bp hypomethylated DMR (T1000_0179) located in an intron at chr3:71,478,001-71,479,000 coordinates. FOXP1 is a transcription factor known to promote CSC characteristics in ovarian cancer, in which its overexpression induces the upregulation of ABCG2, OCT4, NANOG, and SOX2 expression (Choi *et al.*, 2016). Additionally, FOXP1 is pro-oncogenic, promoting breast cancer cell proliferation and tumor growth, and has been linked to breast cancer stem cell maintenance (Chiang *et al.*, 2017; Jing *et al.*, 2021). The upregulation of these 3 genes due to hypomethylated intronic regions in CSCs compared to non-CSCs suggests that they may play a role in the dedifferentiation process through changes in their methylation landscape.

Before investigating further methylation changes and their associated genes, methylation levels between subpopulations have to be verified by a region-specific method to validate the differences identified. Hence, a few DMRs have to be selected among the list of 35 DMRs for the validation step. To select DMRs, one additional parameter was used to further reduce the list: the number of CpG sites covered in all samples to increase the robustness of methylation differences. Hence, by looking at the 35 DMRs with at least 5 CpG sites covered in all samples, 9 DMRs were filtered. Among those 9 DMRs, 5 are associated with interesting genes: ① fascin actin-bundling protein 1 (FSCN1), ② cholinergic receptor nicotinic alpha 6 subunit (CHRNA6), ③ cadherin 7 (CDH7), ④ CD9, and ⑤ protein kinase CAMP-dependent type I regulatory subunit beta (PRKAR1B) genes.

① The FSCN1 gene is located at chr7:5,632,439-5,646,286 coordinates and its associated hypomethylated DMR corresponds to a 328 bp CGI (CGIR_0005) located at chr7:5,647,657-5,647,984 coordinates, 1,370 bp upstream of the FSCN1 gene. FSCN1 promotes EMT in various cancers, including ovarian cancer, squamous cell carcinoma, and lung cancer (Li *et al.*, 2017a; Pan *et al.*, 2017; Wang *et al.*, 2017b). It is a downstream effector of SNAI2, and its expression is significantly elevated after the induction of TGF- β expression (Keshamouni *et al.*, 2009; Wang *et al.*, 2017b). In addition, FSCN1 upregulation has been associated with the promotion of brain metastasis of breast cancer cells, *via* the upregulation of SOX2 expression (Xiao *et al.*, 2020).

② The CHRNA6 gene is located at chr8:42,607,763-42,651,535 coordinates and its associated hypermethylated DMR corresponds to a 17,431 bp methylation segment (SEGR_0040) located at chr8:42,623,718-42,641,148 coordinates, and covers mostly an intron of the CHRNA6 gene. CHRNA6 is a subunit of the neuronal nicotinic acetylcholine receptors, that mediate dopaminergic neurotransmission activated by acetylcholine and nicotine; it is therefore associated with tobacco carcinogenesis in esophageal squamous cell carcinoma and lung cancer (Song *et al.*, 2015; Szymanowska-Narloch *et al.*, 2013).

③ The CDH7 gene is located at chr18:63,417,488-63,548,638 coordinates and its associ-

ated hypermethylated DMR corresponds to a 51 bp methylation segment (SEGR_0087) located at chr18:63,418,129-63,418,179 coordinates, in the first intron of the CDH7 gene and inside a CGI. CDH7 is part of the cadherin family — calcium-dependent cell-cell adhesion molecules — and has been associated with the promotion of Hh signaling in neural cells (Kawano *et al.*, 2017).

④ The CD9 gene is located at chr12:6,308,881-6,347,425 coordinates and its associated hypomethylated DMR corresponds to a 1,001 bp region surrounding its TSS (STSS_0055) located at chr12:6,308,381-6,309,381 coordinates. CD9 is a transmembrane protein involved in adhesion, and signal transduction. In breast cancer, CD9 promotes chemoresistance, migration, invasion, and bone metastasis (Kischel *et al.*, 2012; Rappa *et al.*, 2015; Ullah *et al.*, 2019). Moreover, it promotes stem cell properties of cancer cells, and as it is overexpressed in CSCs, it can be used as a CSC marker in leukemia, glioblastoma, and ovarian cancer (Liu *et al.*, 2021c; Nagare *et al.*, 2020; Nishida *et al.*, 2009; Podergajs *et al.*, 2015; Yamazaki *et al.*, 2011).

⑤ The PRKAR1B gene is located at chr7:588,834-767,287 coordinates and its associated hypomethylated DMR corresponds to a 1,000 bp region (T1000_0372) located at chr7:579,001-580,000 coordinates, 8,833 bp downstream the PRKAR1B gene. The PRKAR1B gene codes for a regulatory subunit of the protein kinase A (PKA, cAMP-dependent protein kinase) but also produces a circular RNA — stable RNA molecule in a closed-loop structure — associated with cancer progression, migration, and invasion in liver cancer, and osteosarcoma (Feng *et al.*, 2021; Liu *et al.*, 2021a).

These 5 genes could be regulated by DNA methylation changes occurring at the identified differentially methylated regions between subpopulations. For the FSCN1 gene, the 1,370 bp upstream differentially methylated CGI in the promoter region is a promising lead in deciphering the methylation regulation of FSCN1 expression and its involvement in the CSC plasticity. Same for the CD9 gene, correlations between its increased expression in CSCs and the hypomethylation of the region surrounding its TSS could be drawn. For the three other genes, as the gene body methylation can alter gene expression, they can also be regulated by DNA methylation changes and implicated in the phenotypic switch (Jjingo *et al.*, 2012; Yang *et al.*, 2014).

The identification of methylation differences based on RRBS methylation analysis was used in this study as the first step to prospect for new genes and mechanisms contributing to the dedifferentiation into CSC. After finding these differences, it is necessary to validate them with a more precise and region-specific technique for the same populations. For example, the bisulfite sequencing PCR (BSP) assay is one of the most convenient and accessible method to quantify methylation levels and validate methylation differences at a specific region. Once methylation differences are confirmed, the evaluation of the effect of DNA methylation regulation on associated gene expressions can be carried out. Indeed, the expression analysis of associated genes can be performed to correlate methylation level changes with expression level changes. Then, the study of the involvement of these validated genes in the radio-induced dedifferentiation can be done by inhibition or overexpression approaches,

similar to the one used to study the effect of the DNMTs inhibition on dedifferentiation.

Also, live tracking approaches with fluorescent reporters to monitor target gene expressions after the induction of dedifferentiation can be carried out to get insights into spatio-temporal dynamics leading to the reacquisition of stem cell properties. As the radio-induced dedifferentiation is a rare and asynchronous event, inhibition or overexpression approaches on the entire population at fixed time points cannot provide information on the sequential order of mechanisms driving the conversion into CSC. Only single-cell live tracking approaches would enable the detection of events prior to the phenotypic reprogramming, allowing to distinguish causes from consequences of the change of state.

Lastly, methylation differences can also be correlated with other epigenetic mechanisms, such as histones post-translational modifications (PTMs), to identify, for example, segments of bivalent chromatin possibly involved in the rapid phenotypic conversion (Chaffer *et al.*, 2013; Suvà *et al.*, 2013; Wainwright and Scaffidi, 2017). A chromatin immunoprecipitation sequencing (ChIP-seq) of H3K4me3 marks, associated with transcriptional activation, and of H3K27me3 marks, associated with transcriptional repression, has been performed on the same populations. ChIP-seq data still need to be analyzed and cross-referenced with RRBS data to potentially identify new epigenetically regulated genes throughout the radio-induced dedifferentiation.

To conclude, this analysis of global methylation resulted in the identification of methylation changes occurring during the radio-induced phenotypic conversion of non-CSCs into CSCs. These differences still need to be validated, by BSP assays for example. Then, the correlation between these validated methylation differences and expression changes, as well as the evaluation of the involvement of these genes in dedifferentiation, will provide new insights into genes and mechanisms required for the CSC plasticity. Targeting those newly identified candidate genes would prevent the radio-induced conversion of cells into a stem-like phenotype, reducing the tumor enrichment in highly resistant CSCs following radiation exposure, and thereby lowering the cancer resistance to radiotherapy.

2 THE ABSP "ANALYSIS OF BISULFITE SEQUENCING PCR" TOOL: USAGE, LIMITATIONS, AND IMPROVEMENTS

As the project unfolded, methylation differences obtained from RRBS analysis needed to be validated. Therefore, the BSP technique was chosen as it is one of the most convenient and accessible way to quantify locus-specific methylation levels. Moreover, the direct-BSP approach was selected for preliminary validation as it requires fewer sequencing runs and therefore it is less expensive than cloning-BSP to get a first insight. However, difficulties in the analysis of sequencing data to get methylation percentages emerged after a first test of direct-BSP. Indeed, the *ESME* (Epigenetic Sequencing Methylation analysis) software, developed by Lewin *et al.* (2004), was the only available tool to compute methylation percentages based on fluorescence intensities from four dye trace sequencing data. However, this tool requires proficiency in a Linux operating system for its installation

and utilization, which is not common for biologists and lacks a user-friendly interface. Additionally, the *ESME* software can only calculate methylation levels and does not provide any further functionalities such as data visualization or comparative analysis. To complement the *ESME*-mediated direct-BSP analysis, another tool can be used for data visualization: the *Methylation plotter* web tool developed by [Mallona et al. \(2014\)](#). This tool can generate graphical representations of methylation data such as lollipop-style plots and methylation profile plots but is limited in terms of statistical analysis.

Hence, the aim was to create an analysis workflow combining functionalities of *ESME* and *Methylation plotter*, and even beyond, to analyze direct-BSP data especially, but also cloning-BSP data for an analytic process continuity between both approaches. Moreover, the emphasis was placed on the complete procedure automatization to maximize user accessibility and provide a ready-to-use solution for researchers. Therefore, the Analysis of Bisulfite Sequencing PCR (ABSP) tool was developed to meet those needs.

In terms of application, studies choosing the direct-BSP approach are currently using the *ESME* software, and their common element is the relatively large amount of samples analyzed. Within the last couple of years, the *ESME* software have been used in these publications (non-exhaustive list): [Achenbach et al. \(2022\)](#); [Carvalho et al. \(2022\)](#); [Heseding et al. \(2022\)](#); [Liu et al. \(2022\)](#); and [Chenarani et al. \(2021\)](#); [Hartung et al. \(2021\)](#); [Liu et al. \(2021b\)](#); [Pfisterer et al. \(2021\)](#); [Schiele et al. \(2021\)](#); [Velásquez et al. \(2021\)](#). As an example, in [Ismail et al. \(2020\)](#), the authors studied the methylation of the PTPRG tumor suppressor gene in chronic myeloid leukemia (CML) patients presenting a high resistance to the imatinib mesylate tyrosine kinase inhibitor treatment. The blood from 26 CML patients and 6 healthy controls was analyzed by direct-BSP assays to evaluate methylation levels of two genomic regions, covering the promoter (25 CpG sites, 321 bp amplicon) and intron-1 (26 CpG sites, 218 bp amplicon) of the PTPRG gene ([Ismail et al., 2020](#)). The analysis of sequencing data was carried out by using the *ESME* software and the *Methylation plotter* web tool. This example illustrates the type of application for which the ABSP tool can be effective.

Even if the direct-BSP approach is not as commonly used as the cloning-BSP, and that this technique can be substituted for more sensitive ones such as pyrosequencing or even targeted bisulfite sequencing coupled with NGS, direct-BSP experiments are still frequently used in recent studies. More importantly, the great accessibility and affordability of the direct-BSP technique is a valuable asset either to get preliminary results, thus minimizing the time and cost before transitioning to more advanced expensive assays, or to confirm region-specific results obtained by genome-wide analyses such as RRBS/WGBS.

Several improvements can still be implemented in the ABSP software to offer additional features. First, the development process of ABSP was centered around the methylation at CpG sites for simplicity and efficiency reasons. Therefore, ABSP is more suited for the study of DNA methylation in mammals. To investigate DNA methylation in plant models for instance, the CHG and CHH contexts

have to be considered in addition to the CG (CpG) context (Chan *et al.*, 2005; Lister *et al.*, 2008; Zhang *et al.*, 2006). The implementation of this feature can be done by adding the search for CHG and CHH contexts on the reference DNA sequence in the “Find CpG coordinates” step of the individual analysis. Hence, it would require either a new option in the input panel of the individual analysis within the interface or a link between the reference genome and the choice of methylation sites, to automatically unlock CHG and CHH contexts when a plant genome is selected as a reference. Additionally, the app and reports contents have to be adjusted to not only mention CpG methylation but to include the different context possibilities. Such an additional feature could be implemented as part of a 2.0 version of the ABSP software.

Then, inspired by the *Methylation plotter* app interface generating interactive plots with a few plotting options, the same principle could be applied to ABSP within its app. Indeed, the generation of a report by the grouped analysis is not the most optimized method to create, display and save plots as it does not offer many parameters to dynamically adjust plots. Indeed, the grouped analysis report is mainly composed of a list of generated plots with limited fixed options, that are non-editable for the user after the report generation. Hence, complementary to the current grouped analysis report, a new tab named “plot generator” could be created to dynamically generate plots of grouped samples directly within the ABSP app as an interactive dashboard. This “plot generator” tab would increase readability and would offer more flexibility when it comes to visualizing results, as it would provide a wide range of options for a customized and dynamic plot generation. In practice, similarly to the grouped analysis, methylation data files from individual analyses could directly be retrieved through the app by selecting the corresponding input folder and sequence name. The data grouping step within the grouped analysis can be transferred into the app, to generate and store the full methylation data table required as an input to create plots. A few additional parameters and options (e.g. plot width and height, font sizes, plot title...) could be added within the different plot functions and linked to input fields within the app to be available for the user. Although the existing data grouping step and plot functions from the grouped analysis can easily be transferred into the app, the addition of several parameters coupled with their corresponding input entries in the app is a more tedious process and would necessarily lead to a major update of ABSP.

To conclude, ABSP provides a complete analytic process of both BSP approaches, it is fully automated, user-friendly, and flexible, and therefore it facilitates the analysis of BSP experiments for biologists. The BSP assay paired with its automated analysis by ABSP will definitely help researchers to rapidly evaluate region-specific DNA methylation levels in order to verify their experimental hypothesis.

BIBLIOGRAPHY

A

- Abdouh, M., Facchino, S., Chatoo, W., Balasingam, V., Ferreira, J., & Bernier, G. (2009). BMI1 Sustains Human Glioblastoma Multiforme Stem Cell Renewal. *Journal of Neuroscience*, *29*(28), 8884–8896. <https://doi.org/10.1523/jneurosci.0968-09.2009>
- Abubaker, K., Latifi, A., Luwor, R., Nazaretian, S., Zhu, H., Quinn, M. A., Thompson, E. W., Findlay, J. K., & Ahmed, N. (2013). Short-term single treatment of chemotherapy results in the enrichment of ovarian cancer stem cell-like cells leading to an increased tumor burden. *Molecular Cancer*, *12*(1), 24–24. <https://doi.org/10.1186/1476-4598-12-24>
- Achenbach, J., Rhein, M., Glahn, A., Frieling, H., & Karst, M. (2022). Leptin promoter methylation in female patients with painful multisomatoform disorder and chronic widespread pain. *Clinical Epigenetics*, *14*(1), 13. <https://doi.org/10.1186/s13148-022-01235-5>
- Aggerholm, A., Guldberg, P., Hokland, M., & Hokland, P. (1999). Extensive intra- and interindividual heterogeneity of p15INK4B methylation in acute myeloid leukemia. *Cancer research*, *59*(2), 436–41.
- Ahmed, M. A. H., Aleskandarany, M. A., Rakha, E. A., Moustafa, R. Z. A., Benhasouna, A., Nolan, C., Green, A. R., Ilyas, M., & Ellis, I. O. (2012). A CD44-/CD24+ phenotype is a poor prognostic marker in early invasive breast cancer. *Breast Cancer Research and Treatment*, *133*(3), 979–995. <https://doi.org/10.1007/s10549-011-1865-8>
- Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B., Nusbaum, C., & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2), R18–R18. <https://doi.org/10.1186/gb-2011-12-2-r18>
- Ajani, J. A., Song, S., Hochster, H. S., & Steinberg, I. B. (2015). Cancer Stem Cells: The Promise and the Potential. *Seminars in Oncology*, *42*, S3–S17. <https://doi.org/10.1053/j.seminoncol.2015.01.001>
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), R87. <https://doi.org/10.1186/gb-2012-13-10-r87>
- Akika, R., Awada, Z., Mogharbil, N., & Zgheib, N. K. (2017). Region of interest methylation analysis: a comparison of MSP with MS-HRM and direct BSP. *Molecular Biology Reports*, *44*(3), 295–305. <https://doi.org/10.1007/s11033-017-4110-7>
- Albino, D., Civenni, G., Dallavalle, C., Roos, M., Jahns, H., Curti, L., Rossi, S., Pinton, S., D'Ambrosio, G., Sessa, F., Hall, J., Catapano, C. V., & Carbone, G. M. (2016). Activation of the Lin28/let-7 Axis by Loss of ESE3/EHF Promotes a Tumorigenic and Stem-like Phenotype in Prostate Cancer. *Cancer Research*, *76*(12), 3629–3643. <https://doi.org/10.1158/0008-5472.can-15-2665>
- Alfarouk, K. O., Stock, C.-M., Taylor, S., Walsh, M., Muddathir, A. K., Verduzco, D., Bashir, A. H. H., Mohammed, O. Y., Elhasan, G. O., Harguindey, S., Reshkin, S. J., Ibrahim, M. E., & Rauch, C. (2015). Resistance to cancer chemotherapy: failure in drug response from ADME to P-gp. *Cancer Cell International*, *15*(1), 71. <https://doi.org/10.1186/s12935-015-0221-1>
- Alhabbab, R. Y. (2020). Targeting Cancer Stem Cells by Genetically Engineered Chimeric Antigen Receptor T Cells. *Frontiers in Genetics*, *11*, 312. <https://doi.org/10.3389/fgene.2020.00312>
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, *100*(7), 3983–3988. <https://doi.org/10.1073/pnas.0530291100>
- Ali, S., Rasool, M., Chaoudhry, H., Pushparaj, P. N., Jha, P., Hafiz, A., Mahfooz, M., Sami, G. A., Kamal, M. A., Bashir, S., Ali, A., & Jamal, M. S. (2016). Molecular mechanisms and mode of tamoxifen resistance in breast cancer. *Bioinformation*, *12*(3), 135–139. <https://doi.org/10.6026/97320630012135>
- Allison, K. H., Hammond, M. E. H., Dowsett, M., McKernin, S. E., Carey, L. A., Fitzgibbons, P. L., Hayes, D. F., Lakhani, S. R., Chavez-MacGregor, M., Perlmutter, J., Perou, C. M., Regan, M. M., Rimm, D. L., Symmans, W. F., Torlakovic, E. E., Varella, L., Viale, G., Weisberg, T. F., McShane, L. M., & Wolff, A. C. (2020). Estrogen and Progesterone Receptor Testing in Breast Cancer: ASCO/CAP Guideline Update. *Journal of Clinical Oncology*, *38*(12), 1346–1366. <https://doi.org/10.1200/jco.19.02309>
- Al-Mayah, A. H. J., Irons, S. L., Pink, R. C., Carter, D. R. F., & Kadhim, M. A. (2012). Possible Role of Exosomes Containing RNA in Mediating Nontargeted Effect of Ionizing Radiation. *Radiation Research*, *177*(5), 539–545. <https://doi.org/10.1667/rr2868.1>
- Alvarez, A. A., Field, M., Bushnev, S., Longo, M. S., & Sugaya, K. (2015). The Effects of Histone Deacetylase Inhibitors on Glioblastoma-Derived Stem Cells. *Journal of Molecular Neuroscience*, *55*(1), 7–20. <https://doi.org/10.1007/s12031-014-0329-0>
- Anbazhagan, R., Herman, J. G., Enika, K., & Gabrielson, E. (2001). Spreadsheet-Based Program for the Analysis of DNA Methylation. *BioTechniques*, *30*(1), 110–114. <https://doi.org/10.2144/01301bc01>

- Andersson, E. R., & Lendahl, U. (2014). Therapeutic modulation of Notch signalling — are we there yet? *Nature Reviews Drug Discovery*, 13(5), 357–378. <https://doi.org/10.1038/nrd4252>
- Andriani, F., Bertolini, G., Facchinetti, F., Baldoli, E., Moro, M., Casalini, P., Caserini, R., Milione, M., Leone, G., Pelosi, G., Pastorino, U., Sozzi, G., & Roz, L. (2016). Conversion to stem-cell state in response to microenvironmental cues is regulated by balance between epithelial and mesenchymal features in lung cancer cells. *Molecular Oncology*, 10(2), 253–271. <https://doi.org/10.1016/j.molonc.2015.10.002>
- Angeloni, A., & Bogdanovic, O. (2019). Enhancer DNA methylation: implications for gene regulation. *Essays in Biochemistry*, 63(6), 707–715. <https://doi.org/10.1042/ebc20190030>
- Antwih, D. A., Gabbara, K. M., Lancaster, W. D., Ruden, D. M., & Zielske, S. P. (2013). Radiation-induced epigenetic DNA methylation modification of radiation-response pathways. *Epigenetics*, 8(8), 839–848. <https://doi.org/10.4161/epi.25498>
- Aponte, P. M., & Caicedo, A. (2017). Stemness in Cancer: Stem Cells, Cancer Stem Cells, and Their Microenvironment. *Stem Cells International*, 2017, 5619472. <https://doi.org/10.1155/2017/5619472>
- Applied Biosystems Genetic Analysis Data File Format. (2006). https://projects.nfstc.org/workshops/resources/articles/ABIF%5C_File%5C_Format.pdf
- Aran, D., & Hellman, A. (2013). DNA Methylation of Transcriptional Enhancers and Cancer Predisposition. *Cell*, 154(1), 11–13. <https://doi.org/10.1016/j.cell.2013.06.018>
- Arand, J., Spieler, D., Karius, T., Branco, M. R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., & Walter, J. (2012). In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genetics*, 8(6), e1002750. <https://doi.org/10.1371/journal.pgen.1002750>
- Arányi, T., & Tusnády, G. E. (2007). PCR Primer Design. *Methods in Molecular Biology™*, 402, 385–402. https://doi.org/10.1007/978-1-59745-528-2_20
- Arányi, T., Váradi, A., Simon, I., & Tusnády, G. E. (2006). The BiSearch web server. *BMC Bioinformatics*, 7(1), 431–431. <https://doi.org/10.1186/1471-2105-7-431>
- Arciero, C. A., Diehl, A. H., Liu, Y., Sun, Q., Gillespie, T., Li, X., & Subhedar, P. (2020). Triple-negative apocrine carcinoma: A rare pathologic subtype with a better prognosis than other triple-negative breast cancers. *Journal of Surgical Oncology*, 122(6), 1232–1239. <https://doi.org/10.1002/jso.26129>
- Armstrong, C. A., Jones, G. D., Anderson, R., Iyer, P., Narayanan, D., Sandhu, J., Singh, R., Talbot, C. J., & Tufarelli, C. (2012). DNMTs are required for delayed genome instability caused by radiation. *Epigenetics*, 7(8), 892–902. <https://doi.org/10.4161/epi.21094>
- Arnold, C., Mangesius, J., Skvortsova, I.-I., & Ganswindt, U. (2020a). The Role of Cancer Stem Cells in Radiation Resistance. *Frontiers in oncology*, 10, 164. <https://doi.org/10.3389/fonc.2020.00164>
- Arnold, K. M., Opendaker, L. M., Flynn, N. J., Appeah, D., & Sims-Mourtada, J. (2020b). Radiation induces an inflammatory response that results in STAT3-dependent changes in cellular plasticity and radioresistance of breast cancer stem-like cells. *International Journal of Radiation Biology*, 1–14. <https://doi.org/10.1080/09553002.2020.1705423>
- Ashapkin, V. V., Kutueva, L. I., & Vanyushin, B. F. (2020). Clinical and Preclinical Models for Maximizing Healthspan, Methods and Protocols. *Methods in Molecular Biology*, 2138, 297–312. https://doi.org/10.1007/978-1-0716-0471-7_21
- Ashrafizadeh, M., Farhood, B., Musa, A. E., Taeb, S., & Najafi, M. (2020). The interactions and communications in tumor resistance to radiotherapy: Therapy perspectives. *International Immunopharmacology*, 87, 106807. <https://doi.org/10.1016/j.intimp.2020.106807>
- Auffinger, B., Tobias, A. L., Han, Y., Lee, G., Guo, D., Dey, M., Lesniak, M. S., & Ahmed, A. U. (2014). Conversion of differentiated cancer cells into cancer stem-like cells in a glioblastoma model after primary chemotherapy. *Cell Death & Differentiation*, 21(7), 1119–1131. <https://doi.org/10.1038/cdd.2014.31>

B

- Baba, T., Convery, P. A., Matsumura, N., Whitaker, R. S., Kondoh, E., Perry, T., Huang, Z., Bentley, R. C., Mori, S., Fujii, S., Marks, J. R., Berchuck, A., & Murphy, S. K. (2009). Epigenetic regulation of CD133 and tumorigenicity of CD133+ ovarian cancer cells. *Oncogene*, 28(2), 209–218. <https://doi.org/10.1038/onc.2008.374>
- Bachelard-Cascales, E., Chapellier, M., Delay, E., Pochon, G., Voeltzel, T., Puisieux, A., Fromentel, C. C. d., & Maguer-Satta, V. (2010). The CD10 Enzyme Is a Key Player to Identify and Regulate Human Mammary Stem Cells. *STEM CELLS*, 28(6), 1081–1088. <https://doi.org/10.1002/stem.435>
- Bae, J.-H., Kim, J.-G., Heo, K., Yang, K., Kim, T.-O., & Yi, J. M. (2015). Identification of radiation-induced aberrant hypomethylation in colon cancer. *BMC Genomics*, 16(1), 56. <https://doi.org/10.1186/s12864-015-1229-6>
- Bailleul, J. (2018). *Etude des mécanismes impliqués dans la reprogrammation de cellules cancéreuses non-souches en cellules souches cancéreuses induite par les radiations ionisantes dans le cancer du sein* (Doctoral dissertation).

- Ball, M. P., Li, J. B., Gao, Y., Lee, J.-H., LeProust, E., Park, I.-H., Xie, B., Daley, G. Q., & Church, G. M. (2009). Targeted and genome-scale methylomics reveals gene body signatures in human cell lines. *Nature biotechnology*, 27(4), 361–368. <https://doi.org/10.1038/nbt.1533>
- Bandhavkar, S. (2016). Cancer stem cells: a metastasizing menace! *Cancer Medicine*, 5(4), 649–655. <https://doi.org/10.1002/cam4.629>
- Bao, S., Wu, Q., McLendon, R. E., Hao, Y., Shi, Q., Hjelmeland, A. B., Dewhirst, M. W., Bigner, D. D., & Rich, J. N. (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature*, 444(7120), 756–760. <https://doi.org/10.1038/nature05236>
- Bapat, S. A. (2013). Epigenetics: Development and Disease. *Subcellular Biochemistry*, 61, 419–434. https://doi.org/10.1007/978-94-007-4525-4_18
- Bar, E. E., Lin, A., Mahairaki, V., Matsui, W., & Eberhart, C. G. (2010). Hypoxia Increases the Expression of Stem-Cell Markers and Promotes Clonogenicity in Glioblastoma Neurospheres. *The American Journal of Pathology*, 177(3), 1491–1502. <https://doi.org/10.2353/ajpath.2010.091021>
- Barlow, D. P., & Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, 6(2), a018382. <https://doi.org/10.1101/cshperspect.a018382>
- Baskar, R., Lee, K. A., Yeo, R., & Yeoh, K.-W. (2012). Cancer and Radiation Therapy: Current Advances and Future Directions. *International Journal of Medical Sciences*, 9(3), 193–199. <https://doi.org/10.7150/ijms.3635>
- Battle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nature Medicine*, 23(10), 1124–1134. <https://doi.org/10.1038/nm.4409>
- Baubec, T., Ivánek, R., Lienert, F., & Schübeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell*, 153(2), 480–492. <https://doi.org/10.1016/j.cell.2013.03.011>
- Baumer, A., Wiedemann, U., Hergersberg, M., & Schinzel, A. (2001). A novel MSP/DHPLC method for the investigation of the methylation status of imprinted genes enables the molecular detection of low cell mosaicisms. *Human Mutation*, 17(5), 423–430. <https://doi.org/10.1002/humu.1118>
- Baylin, S. B., & Jones, P. A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harbor Perspectives in Biology*, 8(9), a019505. <https://doi.org/10.1101/cshperspect.a019505>
- Begicevic, R.-R., & Falasca, M. (2017). ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. *International Journal of Molecular Sciences*, 18(11), 2362. <https://doi.org/10.3390/ijms18112362>
- Behbod, F., & Vivanco, M. d. M. (2015). Mammary Stem Cells, Methods and Protocols. *Methods in Molecular Biology*, 1293, 73–81. https://doi.org/10.1007/978-1-4939-2519-3_4
- Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood. Education and Practice Edition*, 98(6), 236–238. <https://doi.org/10.1136/archdischild-2013-304340>
- Ben-Porath, I., Thomson, M. W., Carey, V. J., Ge, R., Bell, G. W., Regev, A., & Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genetics*, 40(5), 499–507. <https://doi.org/10.1038/ng.127>
- Bentzen, S. M., Agrawal, R. K., Aird, E. G. A., Barrett, J. M., Barrett-Lee, P. J., Bentzen, S. M., Bliss, J. M., Brown, J., Dewar, J. A., Dobbs, H. J., Haviland, J. S., Hoskin, P. J., Hopwood, P., Lawton, P. A., Magee, B. J., Mills, J., Morgan, D. A. L., Owen, J. R., Simmons, S., . . . Yarnold, J. R. (2008a). The UK Standardisation of Breast Radiotherapy (START) Trial B of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *The Lancet*, 371(9618), 1098–1107. [https://doi.org/10.1016/s0140-6736\(08\)60348-7](https://doi.org/10.1016/s0140-6736(08)60348-7)
- Bentzen, S. M., Agrawal, R. K., Aird, E. G. A., Barrett, J. M., Barrett-Lee, P. J., Bliss, J. M., Brown, J., Dewar, J. A., Dobbs, H. J., Haviland, J. S., Hoskin, P. J., Hopwood, P., Lawton, P. A., Magee, B. J., Mills, J., Morgan, D. A. L., Owen, J. R., Simmons, S., Sumo, G., . . . Yarnold, J. R. (2008b). The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *The Lancet Oncology*, 9(4), 331–341. [https://doi.org/10.1016/s1470-2045\(08\)70077-9](https://doi.org/10.1016/s1470-2045(08)70077-9)
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2), 315–326. <https://doi.org/10.1016/j.cell.2006.02.041>
- Bertheau, P., Lehmann-Che, J., Varna, M., Dumay, A., Poirot, B., Porcher, R., Turpin, E., Plassa, L.-F., Roquancourt, A. d., Boursstyn, E., Cremoux, P. d., Janin, A., Giacchetti, S., Espié, M., & Thé, H. d. (2013). p53 in breast cancer subtypes and new insights into response to chemotherapy. *The Breast*, 22, S27–S29. <https://doi.org/10.1016/j.breast.2013.07.005>
- Bhat, K., Saki, M., Vlashi, E., Cheng, F., Duhachek-Muggy, S., Alli, C., Yu, G., Medina, P., He, L., Damoiseaux, R., Pellegrini, M., Zemke, N. R., Nghiemphu, P. L., Cloughesy, T. F., Liau, L. M., Kornblum, H. I., & Pajonk, F. (2020). The

- dopamine receptor antagonist trifluoperazine prevents phenotype conversion and improves survival in mouse models of glioblastoma. *Proceedings of the National Academy of Sciences*, 117(20), 11085–11096. <https://doi.org/10.1073/pnas.1920154117>
- Bhatia, S., Sharma, J., Bukkapatnam, S., Oweida, A., Lennon, S., Phan, A. V., Milner, D., Uyanga, N., Jimeno, A., Raben, D., Somerset, H., Heasley, L., & Karam, S. D. (2018). Inhibition of EphB4-ephrin-B2 signaling enhances response to Cetuximab-radiation therapy in head and neck cancers. *Clinical Cancer Research*, 24(18), clincanres.0327.2018. <https://doi.org/10.1158/1078-0432.ccr-18-0327>
- Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., & Blau, H. M. (2009). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, 463(7284), 1042–1047. <https://doi.org/10.1038/nature08752>
- Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E. W., Liu, Y., Shin, S., Plaia, T. W., Auerbach, J. M., Arking, D. E., Gonzalez, R., Crook, J., Davidson, B., Schulz, T. C., Robins, A., Khanna, A., Sartipy, P., Hyllner, J., Vanguri, P., . . . Fan, J.-B. (2006). Human embryonic stem cells have a unique epigenetic signature. *Genome Research*, 16(9), 1075–1083. <https://doi.org/10.1101/gr.5319906>
- Bibikova, M., & Fan, J.-B. (2009). DNA Methylation, Methods and Protocols. *Methods in Molecular Biology*, 507, 149–163. https://doi.org/10.1007/978-1-59745-522-0_12
- Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., & Gunderson, K. L. (2009). Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics*, 1(1), 177–200. <https://doi.org/10.2217/epi.09.14>
- Bidan, N., Bailleul-Dubois, J., Duval, J., Winter, M., Denoulet, M., Hannebicque, K., El-Sayed, I. Y., Ginestier, C., Forissier, V., Charafe-Jauffret, E., Macario, M., Matsunaga, Y. T., Meignan, S., Anquez, F., Julien, S., Bonnefond, A., Derhourhi, M., Bourhis, X. L., & Lagadec, C. (2019). Transcriptomic Analysis of Breast Cancer Stem Cells and Development of a pALDH1A1:mNeptune Reporter System for Live Tracking. *PROTEOMICS*, 19(21-22), 1800454. <https://doi.org/10.1002/pmic.201800454>
- Bierie, B., Pierce, S. E., Kroeger, C., Stover, D. G., Pattabiraman, D. R., Thiru, P., Donaher, J. L., Reinhardt, F., Chaffer, C. L., Keckesova, Z., & Weinberg, R. A. (2017). Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proceedings of the National Academy of Sciences*, 114(12), E2337–E2346. <https://doi.org/10.1073/pnas.1618298114>
- Bigas, A., & Porcheri, C. (2018). Molecular Mechanisms of Notch Signaling. *Advances in Experimental Medicine and Biology*, 1066, 235–263. https://doi.org/10.1007/978-3-319-89512-3_12
- Bisson, I., & Prowse, D. M. (2009). WNT signaling regulates self-renewal and differentiation of prostate cancer cells with stem cell characteristics. *Cell Research*, 19(6), 683–697. <https://doi.org/10.1038/cr.2009.43>
- Blanpain, C. (2013). Tracing the cellular origin of cancer. *Nature Cell Biology*, 15(2), 126–134. <https://doi.org/10.1038/ncb2657>
- Blanpain, C., & Fuchs, E. (2014). Plasticity of epithelial stem cells in tissue regeneration. *Science*, 344(6189), 1242281. <https://doi.org/10.1126/science.1242281>
- Bloom, H. J. G., & Richardson, W. W. (1957). Histological Grading and Prognosis in Breast Cancer. *British Journal of Cancer*, 11(3), 359–377. <https://doi.org/10.1038/bjc.1957.43>
- Bock, C., Beerman, I., Lien, W.-H., Smith, Z. D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D. J., & Meissner, A. (2012). DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Molecular Cell*, 47(4), 633–647. <https://doi.org/10.1016/j.molcel.2012.06.019>
- Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., & Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, 21(21), 4067–4068. <https://doi.org/10.1093/bioinformatics/bti652>
- Boelens, M. C., Wu, T. J., Nabet, B. Y., Xu, B., Qiu, Y., Yoon, T., Azzam, D. J., Twyman-Saint Victor, C., Wiemann, B. Z., Ishwaran, H., ter Brugge, P. J., Jonkers, J., Slingerland, J., & Minn, A. J. (2014). Exosome Transfer from Stromal to Breast Cancer Cells Regulates Therapy Resistance Pathways. *Cell*, 159(3), 499–513. <https://doi.org/10.1016/j.cell.2014.09.051>
- Boland, M. J., Nazor, K. L., & Loring, J. F. (2014). Epigenetic Regulation of Pluripotency and Differentiation. *Circulation Research*, 115(2), 311–324. <https://doi.org/10.1161/circresaha.115.301517>
- Bolós, V., Blanco, M., Medina, V., & Oncology, A. G. (2008). Notch signalling in cancer stem cells. www.springerlink.com
- Bonanno, C., Shehi, E., Adlerstein, D., & Makrigiorgos, G. M. (2007). MS-FLAG, a Novel Real-Time Signal Generation Method for Methylation-Specific PCR. *Clinical Chemistry*, 53(12), 2119–2127. <https://doi.org/10.1373/clinchem.2007.094011>
- Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, 3(7), 730–737. <https://doi.org/10.1038/nm0797-730>

- Bonnin, D. A. A., Havrda, M. C., Lee, M. C., Liu, H., Zhang, Z., Nguyen, L. N., Harrington, L. X., Hassanpour, S., Cheng, C., & Israel, M. A. (2017). Secretion-mediated STAT3 activation promotes self-renewal of glioma stem-like cells during hypoxia. *Oncogene*, *37*(8), 1107–1118. <https://doi.org/10.1038/onc.2017.404>
- Bonora, G., Rubbi, L., Morselli, M., Ma, F., Chronis, C., Plath, K., & Pellegrini, M. (2019). DNA methylation estimation using methylation-sensitive restriction enzyme bisulfite sequencing (MREBS). *PLoS ONE*, *14*(4), e0214368. <https://doi.org/10.1371/journal.pone.0214368>
- Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., & Balasubramanian, S. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, *336*(6083), 934–937. <https://doi.org/10.1126/science.1220671>
- Bošković, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Ram, E. V. R., Meshorer, E., & Torres-Padilla, M.-E. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes & Development*, *28*(10), 1042–1047. <https://doi.org/10.1101/gad.238881.114>
- Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, *317*(5845), 1760–1764. <https://doi.org/10.1126/science.1147939>
- Botrugno, O. A., Santoro, F., & Minucci, S. (2009). Histone deacetylase inhibitors as a new weapon in the arsenal of differentiation therapies of cancer. *Cancer Letters*, *280*(2), 134–144. <https://doi.org/10.1016/j.canlet.2009.02.027>
- Boustani, J., Grapin, M., Laurent, P.-A., Apetoh, L., & Mirjole, C. (2019). The 6th R of Radiobiology: Reactivation of Anti-Tumor Immune Response. *Cancers*, *11*(6), 860. <https://doi.org/10.3390/cancers11060860>
- Boyd, V. L., Moody, K. I., Karger, A. E., Livak, K. J., Zon, G., & Burns, J. W. (2006). Methylation-dependent fragment separation: Direct detection of DNA methylation by capillary electrophoresis of PCR products from bisulfite-converted genomic DNA. *Analytical Biochemistry*, *354*(2), 266–273. <https://doi.org/10.1016/j.ab.2006.04.009>
- Boyd, V. L., Moody, K. I., Karger, A. E., Livak, K. J., Zon, G., & Burns, J. W. (2007). Methylation-Dependent Fragment Separation: Novel Analysis Of 5-Methyl Cytosine By Capillary Electrophoresis Of Amplified Dna Using Pcr Incorporation Of Chemically Modified Dctp. *Nucleosides, Nucleotides and Nucleic Acids*, *26*(6-7), 629–634. <https://doi.org/10.1080/15257770701490456>
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., & Young, R. A. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, *122*(6), 947–956. <https://doi.org/10.1016/j.cell.2005.08.020>
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Names, A., Temper, V., Razin, A., & Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation. *Nature*, *371*(6496), 435–438. <https://doi.org/10.1038/371435a0>
- Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing. *PLoS ONE*, *6*(1), e14524. <https://doi.org/10.1371/journal.pone.0014524>
- Brierley, J. D., Gospodarowicz, M. K., & Wittekind, C. (2016). *TNM Classification of Malignant Tumours, 8th Edition* (Wiley-Blackwell, Ed.).
- Bright, S., & Kadhim, M. (2018). The future impacts of non-targeted effects. *International Journal of Radiation Biology*, *94*(8), 1–33. <https://doi.org/10.1080/09553002.2018.1454617>
- Brinkman, A. B., Simmer, F., Ma, K., Kaan, A., Zhu, J., & Stunnenberg, H. G. (2010). Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*, *52*(3), 232–236. <https://doi.org/10.1016/j.ymeth.2010.06.012>
- Britton, K., Eyre, R., Harvey, I., Stemke-Hale, K., Browell, D., Lennard, T., & Meeson, A. (2012). Breast cancer, side population cells and ABCG2 expression. *Cancer Letters*, *323*(1), 97–105. <https://doi.org/10.1016/j.canlet.2012.03.041>
- Bronner, C., Alhosin, M., Hamiche, A., & Mousli, M. (2019). Coordinated Dialogue between UHRF1 and DNMT1 to Ensure Faithful Inheritance of Methylated DNA Patterns. *Genes*, *10*(1), 65. <https://doi.org/10.3390/genes10010065>
- Brunner, A. L., Johnson, D. S., Kim, S. W., Valouev, A., Reddy, T. E., Neff, N. F., Anton, E., Medina, C., Nguyen, L., Chiao, E., Oyulu, C. B., Schroth, G. P., Absher, D. M., Baker, J. C., & Myers, R. M. (2009). Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Research*, *19*(6), 1044–1056. <https://doi.org/10.1101/gr.088773.108>
- Brunt, A. M., Wheatley, D., Yarnold, J., Somaiah, N., Kelly, S., Harnett, A., Coles, C., Goodman, A., Bahl, A., Churn, M., Zotova, R., Sydenham, M., Griffin, C. L., Morden, J. P., Bliss, J. M., & Group, o. b. o. t. F.-F. T. M. (2016). Acute skin toxicity associated with a 1-week schedule of whole breast radiotherapy compared with a standard 3-week regimen delivered in the UK FAST-Forward Trial. *Radiotherapy and Oncology*, *120*(1), 114–118. <https://doi.org/10.1016/j.radonc.2016.02.027>
- Brunt, A. M., Haviland, J. S., Wheatley, D. A., Sydenham, M. A., Alhasso, A., Bloomfield, D. J., Chan, C., Churn, M., Cleator, S., Coles, C. E., Goodman, A., Harnett, A., Hopwood, P., Kirby, A. M., Kirwan, C. C., Morris, C., Nabi, Z., Sawyer,

- E., Somaiah, N., ... Yarnold, J. (2020). Hypofractionated breast radiotherapy for 1 week versus 3 weeks (FAST-Forward): 5-year efficacy and late normal tissue effects results from a multicentre, non-inferiority, randomised, phase 3 trial. *The Lancet*, *395*(10237), 1613–1626. [https://doi.org/10.1016/s0140-6736\(20\)30932-6](https://doi.org/10.1016/s0140-6736(20)30932-6)
- Brunt, A., Haviland, J., Kirby, A., Somaiah, N., Wheatley, D., Bliss, J., & Yarnold, J. (2021). Five-fraction Radiotherapy for Breast Cancer: FAST-Forward to Implementation. *Clinical Oncology*, *33*(7), 430–439. <https://doi.org/10.1016/j.clon.2021.04.016>
- Bug, G., Gül, H., Schwarz, K., Pfeifer, H., Kampfmann, M., Zheng, X., Beissert, T., Boehrer, S., Hoelzer, D., Ottmann, O. G., & Ruthardt, M. (2005). Valproic Acid Stimulates Proliferation and Self-renewal of Hematopoietic Stem Cells. *Cancer Research*, *65*(7), 2537–2541. <https://doi.org/10.1158/0008-5472.can-04-3011>
- Butner, J. D., Dogra, P., Chung, C., Ruiz-Ramírez, J., Nizzero, S., Plodinec, M., Li, X., Pan, P.-Y., Chen, S.-h., Cristini, V., Ozpolat, B., Calin, G. A., & Wang, Z. (2022). Dedifferentiation-mediated stem cell niche maintenance in early-stage ductal carcinoma in situ progression: insights from a multiscale modeling study. *Cell Death & Disease*, *13*(5), 485. <https://doi.org/10.1038/s41419-022-04939-x>

C

- Cabillic, F., & Corlu, A. (2016). Regulation of Transdifferentiation and Retrodifferentiation by Inflammatory Cytokines in Hepatocellular Carcinoma. *Gastroenterology*, *151*(4), 607–615. <https://doi.org/10.1053/j.gastro.2016.06.052>
- Caccese, M., Simonelli, M., Villani, V., Rizzato, S., Ius, T., Pasqualetti, F., Russo, M., Rudà, R., Amoroso, R., Bellu, L., Bertorelle, R., Cavallin, F., Dipasquale, A., Carosi, M., Pizzolitto, S., Cesselli, D., Persico, P., Casini, B., Fassan, M., ... Lombardi, G. (2022). Definition of the Prognostic Role of MGMT Promoter Methylation Value by Pyrosequencing in Newly Diagnosed IDH Wild-Type Glioblastoma Patients Treated with Radiochemotherapy: A Large Multicenter Study. *Cancers*, *14*(10), 2425. <https://doi.org/10.3390/cancers14102425>
- Cai, W.-Y., Wei, T.-Z., Luo, Q.-C., Wu, Q.-W., Liu, Q.-F., Yang, M., Ye, G.-D., Wu, J.-F., Chen, Y.-Y., Sun, G.-B., Liu, Y.-J., Zhao, W.-X., Zhang, Z.-M., & Li, B.-A. (2013). The Wnt- β -catenin pathway represses let-7 microRNA expression through transactivation of Lin28 to augment breast cancer stem cell expansion. *Journal of Cell Science*, *126*(13), 2877–2889. <https://doi.org/10.1242/jcs.123810>
- Cancer Today - Global Cancer Observatory. (n.d.). <https://gco.iarc.fr/today>
- Candiloro, I. L. M., Mikeska, T., & Dobrovic, A. (2017). Assessing alternative base substitutions at primer CpG sites to optimise unbiased PCR amplification of methylated sequences. *Clinical Epigenetics*, *9*(1), 31. <https://doi.org/10.1186/s13148-017-0328-4>
- Cano, A., & Nieto, M. A. (2008). Non-coding RNAs take centre stage in epithelial-to-mesenchymal transition. *Trends in Cell Biology*, *18*(8), 357–359. <https://doi.org/10.1016/j.tcb.2008.05.005>
- Carballo, G. B., Honorato, J. R., Lopes, G. P. F. d., & Spohr, T. C. L. d. S. e. (2018). A highlight on Sonic hedgehog pathway. *Cell Communication and Signaling : CCS*, *16*(1), 11. <https://doi.org/10.1186/s12964-018-0220-7>
- Carnero, A., & Leonart, M. (2016). The hypoxic microenvironment: A determinant of cancer stem cell evolution. *BioEssays*, *38*(S1), S65–S74. <https://doi.org/10.1002/bies.201670911>
- Carruthers, R. D., Ahmed, S. U., Ramachandran, S., Strathdee, K., Kurian, K. M., Hedley, A., Gomez-Roman, N., Kalna, G., Neilson, M. P., Gilmour, L., Stevenson, K. H., Hammond, E. M., & Chalmers, A. J. (2018). Replication stress drives constitutive activation of the DNA damage response and radioresistance in glioblastoma stem-like cells. *Cancer Research*, *78*(17), canres.0569.2018. <https://doi.org/10.1158/0008-5472.can-18-0569>
- Carvalho, E., Dias, A., Sousa, A., Lopes, A. M., Martins, S., Pinto, N., Lemos, C., & Alves-Ferreira, M. (2022). A High Methylation Level of a Novel -284 bp CpG Island in the RAMP1 Gene Promoter Is Potentially Associated with Migraine in Women. *Brain Sciences*, *12*(5), 526. <https://doi.org/10.3390/brainsci12050526>
- Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics*, *10*(5), 295–304. <https://doi.org/10.1038/nrg2540>
- Chaffer, C. L., Brueckmann, I., Scheel, C., Kaestli, A. J., Wiggins, P. A., Rodrigues, L. O., Brooks, M., Reinhardt, F., Su, Y., Polyak, K., Arendt, L. M., Kuperwasser, C., Bierie, B., & Weinberg, R. A. (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proceedings of the National Academy of Sciences*, *108*(19), 7950–7955. <https://doi.org/10.1073/pnas.1102454108>
- Chaffer, C. L., Marjanovic, N. D., Lee, T., Bell, G., Kleer, C. G., Reinhardt, F., D'Alessio, A. C., Young, R. A., & Weinberg, R. A. (2013). Poised Chromatin at the ZEB1 Promoter Enables Breast Cancer Cell Plasticity and Enhances Tumorigenicity. *Cell*, *154*(1), 61–74. <https://doi.org/10.1016/j.cell.2013.06.005>
- Chalouni, C., & Doll, S. (2018). Fate of Antibody-Drug Conjugates in Cancer Cells. *Journal of Experimental & Clinical Cancer Research : CR*, *37*(1), 20. <https://doi.org/10.1186/s13046-017-0667-1>

- Chan, S. W.-L., Henderson, I. R., & Jacobsen, S. E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics*, 6(5), 351–360. <https://doi.org/10.1038/nrg1601>
- Chang, L., Graham, P. H., Hao, J., Ni, J., Bucci, J., Cozzi, P. J., Kearsley, J. H., & Li, Y. (2013). Acquisition of epithelial–mesenchymal transition and cancer stem cell phenotypes is associated with activation of the PI3K/Akt/mTOR pathway in prostate cancer radioresistance. *Cell Death & Disease*, 4(10), e875. <https://doi.org/10.1038/cddis.2013.407>
- Chang, M.-S. (2012). Tamoxifen Resistance in Breast Cancer. *Biomolecules and Therapeutics*, 20(3), 256–267. <https://doi.org/10.4062/biomolther.2012.20.3.256>
- Chargari, C., Deutsch, E., Blanchard, P., Gouy, S., Martelli, H., Guérin, F., Dumas, I., Bossi, A., Morice, P., Viswanathan, A. N., & Haie-Meder, C. (2019). Brachytherapy: An overview for clinicians. *CA: A Cancer Journal for Clinicians*, 69(5), 386–401. <https://doi.org/10.3322/caac.21578>
- Chatterjee, A., Rodger, E. J., Morison, I. M., Eccles, M. R., & Stockwell, P. A. (2017). Oral Biology, Molecular Techniques and Applications (Nicholas, C.K. Seymour Gregory J. and Cullinan Mary P. and Heng, Ed.). *Methods in Molecular Biology*, 1537, 249–277. https://doi.org/10.1007/978-1-4939-6685-1_15
- Chau, W. K., Ip, C. K., Mak, A. S. C., Lai, H.-C., & Wong, A. S. T. (2013). c-Kit mediates chemoresistance and tumor-initiating capacity of ovarian cancer cells through activation of Wnt/ β -catenin–ATP-binding cassette G2 signaling. *Oncogene*, 32(22), 2767–2781. <https://doi.org/10.1038/onc.2012.290>
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14(1), 128–128. <https://doi.org/10.1186/1471-2105-14-128>
- Chen, G. G., Gross, J. A., Lutz, P.-E., Vaillancourt, K., Maussion, G., Bramouille, A., Thérout, J.-F., Gardini, E. S., Ehlert, U., Bourret, G., Masurel, A., Lepage, P., Mechawar, N., Turecki, G., & Ernst, C. (2017a). Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics*, 18(1), 96. <https://doi.org/10.1186/s12864-017-3489-9>
- Chen, M.-T., Sun, H.-F., Zhao, Y., Fu, W.-Y., Yang, L.-P., Gao, S.-P., Li, L.-D., Jiang, H.-I., & Jin, W. (2017b). Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. *Scientific Reports*, 7(1), 9254. <https://doi.org/10.1038/s41598-017-10166-8>
- Chen, T., & Li, E. (2004). Structure and Function of Eukaryotic DNA Methyltransferases. *Current Topics in Developmental Biology*, 60, 55–89. [https://doi.org/10.1016/s0070-2153\(04\)60003-2](https://doi.org/10.1016/s0070-2153(04)60003-2)
- Chen, X., Liao, R., Li, D., & Sun, J. (2017c). Induced cancer stem cells generated by radiochemotherapy and their therapeutic implications. *Oncotarget*, 8(10), 17301–17312. <https://doi.org/10.18632/oncotarget.14230>
- Chen, Z.-X., Mann, J. R., Hsieh, C.-L., Riggs, A. D., & Chédin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *Journal of Cellular Biochemistry*, 95(5), 902–917. <https://doi.org/10.1002/jcb.20447>
- Chen, Z.-X., & Riggs, A. D. (2011). DNA Methylation and Demethylation in Mammals. *The Journal of Biological Chemistry*, 286(21), 18347–18353. <https://doi.org/10.1074/jbc.r110.205286>
- Chenarani, N., Emamjomeh, A., Allahverdi, A., Mirmostafa, S., Afsharinia, M. H., & Zahiri, J. (2021). Bioinformatic tools for DNA methylation and histone modification: A survey. *Genomics*, 113(3), 1098–1113. <https://doi.org/10.1016/j.ygeno.2021.03.004>
- Chhibber, A., & Schroeder, B. G. (2008). Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing. *Analytical Biochemistry*, 377(1), 46–54. <https://doi.org/10.1016/j.ab.2008.02.026>
- Chiang, K., Zielinska, A. E., Shaaban, A. M., Sanchez-Bailon, M. P., Jarrold, J., Clarke, T. L., Zhang, J., Francis, A., Jones, L. J., Smith, S., Barbash, O., Guccione, E., Farnie, G., Smalley, M. J., & Davies, C. C. (2017). PRMT5 Is a Critical Regulator of Breast Cancer Stem Cell Function via Histone Methylation and FOXP1 Expression. *Cell Reports*, 21(12), 3498–3513. <https://doi.org/10.1016/j.celrep.2017.11.096>
- Chiou, S.-H., Wang, M.-L., Chou, Y.-T., Chen, C.-J., Hong, C.-F., Hsieh, W.-J., Chang, H.-T., Chen, Y.-S., Lin, T.-W., Hsu, H.-S., & Wu, C.-W. (2010). Coexpression of Oct4 and Nanog Enhances Malignancy in Lung Adenocarcinoma by Inducing Cancer Stem Cell–Like Properties and Epithelial–Mesenchymal Transdifferentiation. *Cancer Research*, 70(24), 10433–10444. <https://doi.org/10.1158/0008-5472.can-10-2638>
- Chiou, S.-H., Yu, C.-C., Huang, C.-Y., Lin, S.-C., Liu, C.-J., Tsai, T.-H., Chou, S.-H., Chien, C.-S., Ku, H.-H., & Lo, J.-F. (2008). Positive Correlations of Oct-4 and Nanog in Oral Cancer Stem-Like Cells and High-Grade Oral Squamous Cell Carcinoma. *Clinical Cancer Research*, 14(13), 4085–4095. <https://doi.org/10.1158/1078-0432.ccr-07-4404>
- Choi, E. J., Seo, E. J., Kim, D. K., Lee, S. I., Kwon, Y. W., Jang, I. H., Kim, K.-H., Suh, D.-S., & Kim, J. H. (2016). FOXP1 functions as an oncogene in promoting cancer stem cell-like characteristics in ovarian cancer cells. *Oncotarget*, 7(3), 3506–3519. <https://doi.org/10.18632/oncotarget.6510>

- Choi, J.-H., Li, Y., Guo, J., Pei, L., Rauch, T. A., Kramer, R. S., Macmil, S. L., Wiley, G. B., Bennett, L. B., Schnabel, J. L., Taylor, K. H., Kim, S., Xu, D., Sreekumar, A., Pfeifer, G. P., Roe, B. A., Caldwell, C. W., Bhalla, K. N., & Shi, H. (2010). Genome-Wide DNA Methylation Maps in Follicular Lymphoma Cells Determined by Methylation-Enriched Bisulfite Sequencing. *PLoS ONE*, *5*(9), e13020. <https://doi.org/10.1371/journal.pone.0013020>
- Chung, W.-M., Molony, R. D., & Lee, Y.-F. (2021). Non-stem bladder cancer cell-derived extracellular vesicles promote cancer stem cell survival in response to chemotherapy. *Stem Cell Research & Therapy*, *12*(1), 533. <https://doi.org/10.1186/s13287-021-02600-6>
- Church, G. M., & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences*, *81*(7), 1991–1995. <https://doi.org/10.1073/pnas.81.7.1991>
- Chuthapisith, S., Eremin, J., El-Sheemey, M., & Eremin, O. (2010). Breast cancer chemoresistance: Emerging importance of cancer stem cells. *Surgical Oncology*, *19*(1), 27–32. <https://doi.org/10.1016/j.suronc.2009.01.004>
- Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, *78*(1), 273–304. <https://doi.org/10.1146/annurev.biochem.77.062706.153223>
- Clara, J. A., Monge, C., Yang, Y., & Takebe, N. (2020). Targeting signalling pathways and the immune microenvironment of cancer stem cells - a clinical update. *Nature Reviews Clinical Oncology*, *17*(4), 204–232. <https://doi.org/10.1038/s41571-019-0293-2>
- Clark, S. J., Harrison, J., Paul, C. L., & Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Research*, *22*(15), 2990–2997. <https://doi.org/10.1093/nar/22.15.2990>
- Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, V., Godwin, J., Gray, R., Hicks, C., James, S., MacKinnon, E., McGale, P., McHugh, T., Peto, R., Taylor, C., Wang, Y., & (EBCTCG), E. B. C. T. C. G. (2006a). Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *The Lancet*, *366*(9503), 2087–2106. [https://doi.org/10.1016/s0140-6736\(05\)67887-7](https://doi.org/10.1016/s0140-6736(05)67887-7)
- Clarke, M. F., Dick, J. E., Dirks, P. B., Eaves, C. J., Jamieson, C. H., Jones, D. L., Visvader, J., Weissman, I. L., & Wahl, G. M. (2006b). Cancer Stem Cells—Perspectives on Current Status and Future Directions: AACR Workshop on Cancer Stem Cells. *Cancer Research*, *66*(19), 9339–9344. <https://doi.org/10.1158/0008-5472.can-06-3126>
- Clement, V., Sanchez, P., Tribolet, N. d., Radovanovic, I., & Altaba, A. R. i. (2007). HEDGEHOG-GLI1 Signaling Regulates Human Glioma Growth, Cancer Stem Cell Self-Renewal, and Tumorigenicity. *Current Biology*, *17*(2), 165–172. <https://doi.org/10.1016/j.cub.2006.11.033>
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., & Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, *452*(7184), 215–219. <https://doi.org/10.1038/nature06745>
- Colella, S., Shen, L., Baggerly, K., Issa, J.-P., & Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *BioTechniques*, *35*(1), 146–150. <https://doi.org/10.2144/03351md01>
- Colleoni, M., Sun, Z., Price, K. N., Karlsson, P., Forbes, J. F., Thürlimann, B., Gianni, L., Castiglione, M., Gelber, R. D., Coates, A. S., & Goldhirsch, A. (2016). Annual Hazard Rates of Recurrence for Breast Cancer During 24 Years of Follow-Up: Results From the International Breast Cancer Study Group Trials I to V. *Journal of Clinical Oncology*, *34*(9), 927–935. <https://doi.org/10.1200/jco.2015.62.3504>
- Coni, S., Infante, P., & Gulino, A. (2013). Control of stem cells and cancer stem cells by Hedgehog signaling: Pharmacologic clues from pathway dissection. *Biochemical Pharmacology*, *85*(5), 623–628. <https://doi.org/10.1016/j.bcp.2012.11.001>
- Conigliaro, A., Costa, V., Dico, A. L., Saieva, L., Buccheri, S., Dieli, F., Manno, M., Raccosta, S., Mancone, C., Tripodi, M., Leo, G. D., & Alessandro, R. (2015). CD90+ liver cancer cells modulate endothelial cell phenotype through the release of exosomes containing H19 lncRNA. *Molecular Cancer*, *14*(1), 155. <https://doi.org/10.1186/s12943-015-0426-x>
- Conte, P. F., Gennari, A., Landucci, E., & Orlandini, C. (2000). Role of Epirubicin in Advanced Breast Cancer. *Clinical Breast Cancer*, *1*, S46–S51. <https://doi.org/10.3816/cbc.2000.s.009>
- Cortes, J. E., Heidel, F. H., Hellmann, A., Fiedler, W., Smith, B. D., Robak, T., Montesinos, P., Pollyea, D. A., DesJardins, P., Ottmann, O., Ma, W. W., Shaik, M. N., Laird, A. D., Zeremski, M., O'Connell, A., Chan, G., & Heuser, M. (2019). Randomized comparison of low dose cytarabine with or without glasdegib in patients with newly diagnosed acute myeloid leukemia or high-risk myelodysplastic syndrome. *Leukemia*, *33*(2), 379–389. <https://doi.org/10.1038/s41375-018-0312-9>
- Cortes, J. E., Smith, B. D., Wang, E. S., Merchant, A., Oehler, V. G., Arellano, M., DeAngelo, D. J., Pollyea, D. A., Sekeres, M. A., Robak, T., Ma, W. W., Zeremski, M., Shaik, M. N., Laird, A. D., O'Connell, A., Chan, G., & Schroeder, M. A. (2018). Glasdegib in combination with cytarabine and daunorubicin in patients with AML or high-risk MDS: Phase 2 study results. *American Journal of Hematology*, *93*(11), 1301–1310. <https://doi.org/10.1002/ajh.25238>

- Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D. N., Li, Z., Xu, M., Reik, W., Silva, J. C., & Wang, J. (2013). Nanog-dependent function of Tet1 and Tet2 in establishment of pluripotency. *Nature*, *495*(7441), 370–374. <https://doi.org/10.1038/nature11925>
- Costello, J. F., Frühwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomäki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Huang, H.-J. S., Petrelli, N. J., Zhang, X., . . . Plass, C. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature Genetics*, *24*(2), 132–138. <https://doi.org/10.1038/72785>
- Craene, B. D., & Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews Cancer*, *13*(2), 97–110. <https://doi.org/10.1038/nrc3447>
- Crocker, A. K., & Allan, A. L. (2011). Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44+ human breast cancer cells. *Breast Cancer Research and Treatment*, *133*(1), 75–87. <https://doi.org/10.1007/s10549-011-1692-y>
- Crocker, A. K., Goodale, D., Chu, J., Postenka, C., Hedley, B. D., Hess, D. A., & Allan, A. L. (2008). High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. *Journal of Cellular and Molecular Medicine*, *13*(8b), 2236–2252. <https://doi.org/10.1111/j.1582-4934.2008.00455.x>
- Cross, S. H., & Bird, A. P. (1995). CpG islands and genes. *Current Opinion in Genetics & Development*, *5*(3), 309–314. [https://doi.org/10.1016/0959-437x\(95\)80044-1](https://doi.org/10.1016/0959-437x(95)80044-1)
- Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., Ijcken, W. F. v., Grosveld, F., Frommolt, P., Bazzi, H., & Rada-Iglesias, A. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*, *20*(5), 689–705.e9. <https://doi.org/10.1016/j.stem.2017.02.004>

D

- Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, *15*(2), 81–94. <https://doi.org/10.1038/nrclinonc.2017.166>
- Dahl, C., & Guldberg, P. (2003). DNA methylation analysis techniques. *Biogerontology*, *4*(4), 233–250. <https://doi.org/10.1023/a:1025103319328>
- Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J., & Shi, B. (2015). Breast cancer intrinsic subtype classification, clinical use and future trends. *American journal of cancer research*, *5*(10), 2929–43.
- Dai, X., Guo, Y., Hu, Y., Bao, X., Zhu, X., Fu, Q., Zhang, H., Tong, Z., Liu, L., Zheng, Y., Zhao, P., & Fang, W. (2021). Immunotherapy for targeting cancer stem cells in hepatocellular carcinoma. *Theranostics*, *11*(7), 3489–3501. <https://doi.org/10.7150/thno.54648>
- Daniel, C. W., Ome, K. B. D., Young, J. T., Blair, P. B., & Faulkin, L. J. (1968). The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. *Proceedings of the National Academy of Sciences*, *61*(1), 53–60. <https://doi.org/10.1073/pnas.61.1.53>
- Daniel, Y., Lelou, E., Aninat, C., Corlu, A., & Cabillic, F. (2021). Interplay between Metabolism Reprogramming and Epithelial-to-Mesenchymal Transition in Cancer Stem Cells. *Cancers*, *13*(8), 1973. <https://doi.org/10.3390/cancers13081973>
- Danielsson, A., Barreau, K., Kling, T., Tisell, M., & Carén, H. (2020). Accumulation of DNA methylation alterations in paediatric glioma stem cells following fractionated dose irradiation. *Clinical Epigenetics*, *12*(1), 26. <https://doi.org/10.1186/s13148-020-0817-8>
- Darwiche, N. (2020). Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. *American journal of cancer research*, *10*(7), 1954–1978.
- Das, K., & Roychoudhury, A. (2014). Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Frontiers in Environmental Science*, *2*, 53. <https://doi.org/10.3389/fenvs.2014.00053>
- Das, P. K., Pillai, S., Rakib, M. A., Khanam, J. A., Gopalan, V., Lam, A. K. Y., & Islam, F. (2020). Plasticity of Cancer Stem Cell: Origin and Role in Disease Progression and Therapy Resistance. *Stem Cell Reviews and Reports*, *16*(2), 397–412. <https://doi.org/10.1007/s12015-019-09942-y>
- Davis, A. L., Korkaya, H., Ouzounova, M., Malik, F., Kim, G., Quraishi, A. A., Liu, S., Clouthier, S. G., & Wicha, M. S. (2013). Abstract 3736: IL6 feedback loop regulates breast cancer stem cells which mediate trastuzumab resistance. *Cancer Research*, *73*(8_Supplement), 3736–3736. <https://doi.org/10.1158/1538-7445.am2013-3736>
- Dear, S., & Staden, R. (1992). A standard file format for data from DNA sequencing instruments. *DNA Sequence*, *3*(2), 107–110. <https://doi.org/10.3109/10425179209034003>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>

- Debeb, B. G., Lacerda, L., Xu, W., Larson, R., Solley, T., Atkinson, R., Sulman, E. P., Ueno, N. T., Krishnamurthy, S., Reuben, J. M., Buchholz, T. A., & Woodward, W. A. (2012). Histone Deacetylase Inhibitors Stimulate Dedifferentiation of Human Breast Cancer Cells Through WNT/ β -Catenin Signaling. *STEM CELLS*, *30*(11), 2366–2377. <https://doi.org/10.1002/stem.1219>
- Debeb, B. G., Xu, W., Mok, H., Li, L., Robertson, F., Ueno, N. T., Reuben, J., Lucci, A., Cristofanilli, M., & Woodward, W. A. (2010). Differential Radiosensitizing Effect of Valproic Acid in Differentiation Versus Self-Renewal Promoting Culture Conditions. *International Journal of Radiation Oncology*Biophysics*, *76*(3), 889–895. <https://doi.org/10.1016/j.ijrobp.2009.09.052>
- Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., & Walker, P. R. (2005). Sequential DNA Methylation of the Nanog and Oct-4 Upstream Regions in Human NT2 Cells during Neuronal Differentiation. *Journal of Biological Chemistry*, *280*(8), 6257–6260. <https://doi.org/10.1074/jbc.c400479200>
- DeGorter, M., Xia, C., Yang, J., & Kim, R. (2012). Drug Transporters in Drug Efficacy and Toxicity. *Annual Review of Pharmacology and Toxicology*, *52*(1), 249–273. <https://doi.org/10.1146/annurev-pharmtox-010611-134529>
- Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E. M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.-H., Yu, J., Daley, G. Q., Eggan, K., Hochedlinger, K., Thomson, J., Wang, W., Gao, Y., & Zhang, K. (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnology*, *27*(4), 353–360. <https://doi.org/10.1038/nbt.1530>
- Deobagkar, D., Shankar, V., & Deobagkar, D. (1986). Separation of 5-methylcytosine-rich DNA using immobilized antibody. *Enzyme and Microbial Technology*, *8*(2), 97–100. [https://doi.org/10.1016/0141-0229\(86\)90079-7](https://doi.org/10.1016/0141-0229(86)90079-7)
- DeOme, K. B., Faulkin, L. J., Bern, H. A., & Blair, P. B. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer research*, *19*(5), 515–20.
- Dhingra, T., Mittal, K., & Sarma, G. S. (2014). Analytical Techniques for DNA Methylation – An Overview. *Current Pharmaceutical Analysis*, *10*(1), 71–85. <https://doi.org/10.2174/157341291001140102111956>
- Diagenode. (2019). Premium RRBS kit manual. https://www.diagenode.com/files/products/kits/Premium%5C_RRBS%5C_kit%5C_manual.pdf
- Dick, J. E. (2008). Stem cell concepts renew cancer research. *Blood*, *112*(13), 4793–4807. <https://doi.org/10.1182/blood-2008-08-077941>
- Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., & West, A. G. (2010). VEZF1 Elements Mediate Protection from DNA Methylation. *PLoS Genetics*, *6*(1), e1000804. <https://doi.org/10.1371/journal.pgen.1000804>
- Diehn, M., Cho, R. W., & Clarke, M. F. (2009). Therapeutic Implications of the Cancer Stem Cell Hypothesis. *Seminars in Radiation Oncology*, *19*(2), 78–86. <https://doi.org/10.1016/j.semradonc.2008.11.002>
- Ding, Z., Yuan, J., Liang, Y., Wu, J., Gong, H., Ye, Y., Jiang, G., Yin, P., Li, Y., Zhang, G., Yang, C., Guo, J., Chen, Z., Wang, X., Weng, L., & Zou, Y. (2017). Ryanodine Receptor Type 2 Plays a Role in the Development of Cardiac Fibrosis under Mechanical Stretch Through TGF β -1. *International Heart Journal*, *58*(6), 16–572. <https://doi.org/10.1536/ihj.16-572>
- Dobosy, J., & Selker, E. (2001). Emerging connections between DNA methylation and histone acetylation. *Cellular and Molecular Life Sciences CMLS*, *58*(5), 721–727. <https://doi.org/10.1007/pl00000895>
- Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., Miller, J., Schlaeger, T., Daley, G. Q., & Feinberg, A. P. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genetics*, *41*(12), 1350–1353. <https://doi.org/10.1038/ng.471>
- Dombret, H., Seymour, J. F., Butrym, A., Wierzbowska, A., Selleslag, D., Jang, J. H., Kumar, R., Cavenagh, J., Schuh, A. C., Candoni, A., Récher, C., Sandhu, I., Castillo, T. B. d., Al-Ali, H. K., Martinelli, G., Falantes, J., Noppeney, R., Stone, R. M., Minden, M. D., ... Döhner, H. (2015). International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. *Blood*, *126*(3), 291–299. <https://doi.org/10.1182/blood-2015-01-621664>
- Donaldson, A. R., McCarthy, C., Goraya, S., Pederson, H. J., Sturgis, C. D., Grobmyer, S. R., & Calhoun, B. C. (2018). Breast cancer risk associated with atypical hyperplasia and lobular carcinoma in situ initially diagnosed on core-needle biopsy. *Cancer*, *124*(3), 459–465. <https://doi.org/10.1002/cncr.31061>
- Donnarumma, E., Fiore, D., Nappa, M., Roscigno, G., Adamo, A., Iaboni, M., Russo, V., Affinito, A., Puoti, I., Quintavalle, C., Rienzo, A., Piscuoglio, S., Thomas, R., & Condorelli, G. (2017). Cancer-associated fibroblasts release exosomal microRNAs that dictate an aggressive phenotype in breast cancer. *Oncotarget*, *8*(12), 19592–19608. <https://doi.org/10.18632/oncotarget.14752>
- Dontu, G., Al-Hajj, M., Abdallah, W. M., Clarke, M. F., & Wicha, M. S. (2003). Stem cells in normal breast development and breast cancer. *Cell Proliferation*, *36*, 59–72. <https://doi.org/10.1046/j.1365-2184.36.s.1.6.x>

- Dooijeweert, C. v., Deckers, I. A. G., Baas, I. O., Wall, E. v. d., & Diest, P. J. v. (2019). Hormone- and HER2-receptor assessment in 33,046 breast cancer patients: a nationwide comparison of positivity rates between pathology laboratories in the Netherlands. *Breast Cancer Research and Treatment*, *175*(2), 487–497. <https://doi.org/10.1007/s10549-019-05180-5>
- Down, T. A., Rakyán, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Gráf, S., Johnson, N., Herrero, J., Tomazou, E. M., Thorne, N. P., Bäckdahl, L., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Marioni, J. C., Birney, E., Hubbard, T. J. P., . . . Beck, S. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, *26*(7), 779–785. <https://doi.org/10.1038/nbt1414>
- Dubrovská, A., Hartung, A., Bouchez, L. C., Walker, J. R., Reddy, V. A., Cho, C. Y., & Schultz, P. G. (2012). CXCR4 activation maintains a stem cell population in tamoxifen-resistant breast cancer cells through AhR signalling. *British Journal of Cancer*, *107*(1), 43–52. <https://doi.org/10.1038/bjc.2012.105>
- Dunne-Daly, C. F. (1999). Principles of radiotherapy and radiobiology. *Seminars in Oncology Nursing*, *15*(4), 250–259. [https://doi.org/10.1016/s0749-2081\(99\)80054-0](https://doi.org/10.1016/s0749-2081(99)80054-0)
- Dvorakova, M., & Vanek, T. (2016). Histone deacetylase inhibitors for the treatment of cancer stem cells. *MedChemComm*, *7*(12), 2217–2231. <https://doi.org/10.1039/c6md00297h>

E

- Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V., & Laird, P. W. (2000). MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Research*, *28*(8), 32e–. <https://doi.org/10.1093/nar/28.8.e32>
- Easwaran, H., Tsai, H.-C., & Baylin, S. B. (2014). Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Molecular Cell*, *54*(5), 716–727. <https://doi.org/10.1016/j.molcel.2014.05.015>
- Eaves, C. J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood*, *125*(17), 2605–2613. <https://doi.org/10.1182/blood-2014-12-570200>
- Eckert, K. A., & Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *Genome Research*, *1*(1), 17–24. <https://doi.org/10.1101/gr.1.1.17>
- Edwards, J. R., Yarychivska, O., Boulard, M., & Bestor, T. H. (2017). DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin*, *10*(1), 23. <https://doi.org/10.1186/s13072-017-0130-8>
- Eguizabal, C., Montserrat, N., Veiga, A., & Belmonte, J. (2013). Dedifferentiation, Transdifferentiation, and Reprogramming: Future Directions in Regenerative Medicine. *Seminars in Reproductive Medicine*, *31*(01), 082–094. <https://doi.org/10.1055/s-0032-1331802>
- Ehrlich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C. R., Field, J. K., & Boom, D. v. d. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences*, *102*(44), 15785–15790. <https://doi.org/10.1073/pnas.0507816102>
- Ehrlich, M., Gama-Sosa, M. A., Huang, L.-H., Midgett, R. M., Kuo, K. C., McCune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, *10*(8), 2709–2721. <https://doi.org/10.1093/nar/10.8.2709>
- Elston, C., & Ellis, I. (1991). pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, *19*(5), 403–410. <https://doi.org/10.1111/j.1365-2559.1991.tb00229.x>
- Eom, S. H., Wang, J., & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature*, *382*(6588), 278–281. <https://doi.org/10.1038/382278a0>
- Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., & Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature structural & molecular biology*, *15*(11), 1176–1183. <https://doi.org/10.1038/nsmb.1476>
- Eramo, A., Ricci-Vitiani, L., Zeuner, A., Pallini, R., Lotti, F., Sette, G., Pilozzi, E., Larocca, L. M., Peschle, C., & Maria, R. D. (2006). Chemotherapy resistance of glioblastoma stem cells. *Cell Death & Differentiation*, *13*(7), 1238–1241. <https://doi.org/10.1038/sj.cdd.4401872>
- Eroles, P., Bosch, A., Pérez-Fidalgo, J. A., & Lluch, A. (2012). Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews*, *38*(6), 698–707. <https://doi.org/10.1016/j.ctrv.2011.11.005>
- Es, J. H. v., Gijn, M. E. v., Riccio, O., Born, M. v. d., Vooijs, M., Begthel, H., Cozijnsen, M., Robine, S., Winton, D. J., Radtke, F., & Clevers, H. (2005). Notch/ γ -secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature*, *435*(7044), 959–963. <https://doi.org/10.1038/nature03659>

- Escrivá-de-Romaní, S., Arumí, M., Bellet, M., & Saura, C. (2018). HER2-positive breast cancer: Current and new therapeutic strategies. *The Breast*, 39, 80–88. <https://doi.org/10.1016/j.breast.2018.03.006>
- Esposito, A., Criscitiello, C., Salè, E. O., & Curigliano, G. (2014). Optimal adjuvant chemotherapy in breast cancer: selection of agents. *Expert Review of Clinical Pharmacology*, 7(5), 605–611. <https://doi.org/10.1586/17512433.2014.945429>
- Essex, A., Pineda, J., Acharya, G., Xin, H., Evans, J., Iorns, E., Tsui, R., Denis, A., Perfito, N., Errington, T. M., Iorns, E., Tsui, R., Denis, A., Perfito, N., & Errington, T. M. (2019). Replication Study: Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *eLife*, 8, e45426. <https://doi.org/10.7554/elife.45426>
- Estécio, M. R., Yan, P. S., Ibrahim, A. E., Tellez, C. S., Shen, L., Huang, T. H.-M., & Issa, J.-P. J. (2007). High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Research*, 17(10), 1529–1536. <https://doi.org/10.1101/gr.6417007>
- Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I. C., Repasky, E. A., Gabrielson, E., Schutte, M., Baylin, S. B., & Herman, J. G. (2000). Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors. *JNCI: Journal of the National Cancer Institute*, 92(7), 564–569. <https://doi.org/10.1093/jnci/92.7.564>
- Eun, K., Ham, S. W., & Kim, H. (2017). Cancer stem cell heterogeneity: origin and new perspectives on CSC targeting. *BMB Reports*, 50(3), 117–125. <https://doi.org/10.5483/bmbrep.2017.50.3.222>

F

- Falk, S. (2009). Principles of cancer treatment by radiotherapy. *Surgery (Oxford)*, 27(4), 169–172. <https://doi.org/10.1016/j.mpsur.2009.01.008>
- Fallah, Y., Brundage, J., Allegakoen, P., & Shajahan-Haq, A. N. (2017). MYC-Driven Pathways in Breast Cancer Subtypes. *Biomolecules*, 7(3), 53. <https://doi.org/10.3390/biom7030053>
- Fan, H., Zhao, Z., Cheng, Y., Cui, H., Qiao, F., Wang, L., Hu, J., Wu, H., & Song, W. (2016). Genome-wide profiling of DNA methylation reveals preferred sequences of DNMTs in hepatocellular carcinoma cells. *Tumor Biology*, 37(1), 877–885. <https://doi.org/10.1007/s13277-015-3202-z>
- Fan, Y., Liang, Y., Deng, K., Zhang, Z., Zhang, G., Zhang, Y., & Wang, F. (2020). Analysis of DNA methylation profiles during sheep skeletal muscle development using whole-genome bisulfite sequencing. *BMC Genomics*, 21(1), 327. <https://doi.org/10.1186/s12864-020-6751-5>
- Fanelli, G. N., Naccarato, A. G., & Scatena, C. (2020). Recent Advances in Cancer Plasticity: Cellular Mechanisms, Surveillance Strategies, and Therapeutic Optimization. *Frontiers in Oncology*, 10, 569. <https://doi.org/10.3389/fonc.2020.00569>
- Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., MacGrogan, G., Bergh, J., Cameron, D., Goldstein, D., Duss, S., Nicoulaz, A.-L., Brisken, C., Fiche, M., Delorenzi, M., & Iggo, R. (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene*, 24(29), 4660–4671. <https://doi.org/10.1038/sj.onc.1208561>
- Fazekas, A. J., Steeves, R., & Newmaster, S. G. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques*, 48(4), 277–285. <https://doi.org/10.2144/000113369>
- Felice, L. D., Tatarelli, C., Mascolo, M. G., Gregorj, C., Agostini, F., Fiorini, R., Gelmetti, V., Pascale, S., Padula, F., Petrucci, M. T., Arcese, W., & Nervi, C. (2005). Histone Deacetylase Inhibitor Valproic Acid Enhances the Cytokine-Induced Expansion of Human Hematopoietic Stem Cells. *Cancer Research*, 65(4), 1505–1513. <https://doi.org/10.1158/0008-5472.can-04-3063>
- Feng, H., Conneely, K., & Wu, H. (2022). Epigenome-Wide Association Studies, Methods and Protocols. *Methods in Molecular Biology*, 2432, 211–226. https://doi.org/10.1007/978-1-0716-1994-0_16
- Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, 42(8), e69–e69. <https://doi.org/10.1093/nar/gku154>
- Feng, H., & Wu, H. (2019). Differential methylation analysis for bisulfite sequencing using DSS. *Quantitative Biology*, 7(4), 327–334. <https://doi.org/10.1007/s40484-019-0183-8>
- Feng, S., Yao, J., Chen, Y., Geng, P., Zhang, H., Ma, X., Zhao, J., & Yu, X. (2015a). Expression and Functional Role of Reprogramming-Related Long Noncoding RNA (lincRNA-ROR) in Glioma. *Journal of Molecular Neuroscience*, 56(3), 623–630. <https://doi.org/10.1007/s12031-014-0488-z>
- Feng, Z.-h., Zheng, L., Yao, T., Tao, S.-y., Wei, X.-a., Zheng, Z.-y., Zheng, B.-j., Zhang, X.-y., Huang, B., Liu, J.-h., Chen, Y.-l., Shan, Z., Yuan, P.-t., Wang, C.-g., Chen, J., Shen, S.-y., & Zhao, F.-d. (2021). EIF4A3-induced circular RNA PRKAR1B promotes osteosarcoma progression by miR-361-3p-mediated induction of FZD4 expression. *Cell Death & Disease*, 12(11), 1025. <https://doi.org/10.1038/s41419-021-04339-7>

- Feng, Z.-M., Qiu, J., Chen, X.-W., Liao, R.-X., Liao, X.-Y., Zhang, L.-P., Chen, X., Li, Y., Chen, Z.-T., & Sun, J.-G. (2015b). Essential role of miR-200c in regulating self-renewal of breast cancer stem cells and their counterparts of mammary epithelium. *BMC Cancer*, *15*(1), 645. <https://doi.org/10.1186/s12885-015-1655-5>
- Ferguson, A. T., Vertino, P. M., Spitzner, J. R., Baylin, S. B., Muller, M. T., & Davidson, N. E. (1997). Role of Estrogen Receptor Gene Demethylation and DNA Methyltransferase-DNA Adduct Formation in 5-Aza-2'-deoxycytidine-induced Cytotoxicity In Human Breast Cancer Cells. *Journal of Biological Chemistry*, *272*(51), 32260–32266. <https://doi.org/10.1074/jbc.272.51.32260>
- Fillmore, C. M., & Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Research : BCR*, *10*(2), R25–R25. <https://doi.org/10.1186/bcr1982>
- Forn, M., Díez-Villanueva, A., Merlos-Suárez, A., Muñoz, M., Lois, S., Carriò, E., Jordà, M., Bigas, A., Battle, E., & Peinado, M. A. (2015). Overlapping DNA Methylation Dynamics in Mouse Intestinal Cell Differentiation and Early Stages of Malignant Progression. *PLoS ONE*, *10*(5), e0123263. <https://doi.org/10.1371/journal.pone.0123263>
- Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Neste, L. V., Jaenisch, R., & Fan, G. (2008). Promoter CpG Methylation Contributes to ES Cell Gene Regulation in Parallel with Oct4/Nanog, PcG Complex, and Histone H3 K4/K27 Trimethylation. *Cell Stem Cell*, *2*(2), 160–169. <https://doi.org/10.1016/j.stem.2007.12.011>
- Fraga, M. F., Uriol, E., Diego, L. B., Berdasco, M., Esteller, M., Cañal, M. J., & Rodríguez, R. (2002). High-performance capillary electrophoretic method for the quantification of 5-methyl 2'-deoxycytidine in genomic DNA: Application to plant, animal and human cancer tissues. *ELECTROPHORESIS*, *23*(11), 1677–1681. [https://doi.org/10.1002/1522-2683\(200206\)23:11<1677::aid-elps1677>3.0.co;2-z](https://doi.org/10.1002/1522-2683(200206)23:11<1677::aid-elps1677>3.0.co;2-z)
- Franchet, C., Djerroudi, L., Maran-Gonzalez, A., Abramovici, O., Antoine, M., Becette, V., Berghian, A., Blanc-Fournier, C., Brabencova, E., Charafe-Jauffret, E., Chenard, M.-P., Dauplat, M.-M., Delrée, P., Duprez-Paumier, R., Fleury, C., Ghnassia, J.-P., Haudebourg, J., Leroux, A., MacGrogan, G., ... GEFPICS, p. I. (2021). Mise à jour 2021 des recommandations du GEFPICS pour l'évaluation du statut HER2 dans les cancers infiltrants du sein en France. *Annales de Pathologie*, *41*(6), 507–520. <https://doi.org/10.1016/j.annpat.2021.07.014>
- Frank, N. Y., Schatton, T., & Frank, M. H. (2010). The therapeutic promise of the cancer stem cell concept. *Journal of Clinical Investigation*, *120*(1), 41–50. <https://doi.org/10.1172/jci41004>
- French, R., & Pauklin, S. (2021). Epigenetic regulation of cancer stem cell formation and maintenance. *International Journal of Cancer*, *148*(12), 2884–2897. <https://doi.org/10.1002/ijc.33398>
- Frigola, J., Ribas, M., Risques, R.-A., & Peinado, M. A. (2002). Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Research*, *30*(7), e28–e28. <https://doi.org/10.1093/nar/30.7.e28>
- Friso, S., Choi, S.-W., Dolnikowski, G. G., & Selhub, J. (2002). A Method to Assess Genomic DNA Methylation Using High-Performance Liquid Chromatography/Electrospray Ionization Mass Spectrometry. *Analytical Chemistry*, *74*(17), 4526–4531. <https://doi.org/10.1021/ac020050h>
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, *89*(5), 1827–1831. <https://doi.org/10.1073/pnas.89.5.1827>
- Fu, N. Y., Nolan, E., Lindeman, G. J., & Visvader, J. E. (2020). Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiological Reviews*, *100*(2), 489–523. <https://doi.org/10.1152/physrev.00040.2018>
- Fu, N., Lindeman, G. J., & Visvader, J. E. (2014). Chapter Five The Mammary Stem Cell Hierarchy. *Current Topics in Developmental Biology*, *107*, 133–160. <https://doi.org/10.1016/b978-0-12-416022-4.00005-6>
- Fu, Z., Li, G., Li, Z., Wang, Y., Zhao, Y., Zheng, S., Ye, H., Luo, Y., Zhao, X., Wei, L., Liu, Y., Lin, Q., Zhou, Q., & Chen, R. (2017). Endogenous miRNA Sponge LincRNA-ROR promotes proliferation, invasion and stem cell-like phenotype of pancreatic cancer cells. *Cell Death Discovery*, *3*(1), 17004. <https://doi.org/10.1038/cddiscovery.2017.4>
- Fuchs, E., & Chen, T. (2013). A matter of life and death: self-renewal in stem cells. *EMBO reports*, *14*(1), 39–48. <https://doi.org/10.1038/embor.2012.197>
- Fujiwara, S., Kawamoto, T., Kawakami, Y., Koterazawa, Y., Hara, H., Takemori, T., Kitayama, K., Yahiro, S., Kakutani, K., Matsumoto, T., Matsushita, T., Niikura, T., Koyanagi-Aoi, M., Aoi, T., Kuroda, R., & Akisue, T. (2020). Acquisition of cancer stem cell properties in osteosarcoma cells by defined factors. *Stem Cell Research & Therapy*, *11*(1), 429. <https://doi.org/10.1186/s13287-020-01944-9>
- Fulawka, L., Donizy, P., & Halon, A. (2014). Cancer stem cells – the current status of an old concept: literature review and clinical approaches. *Biological Research*, *47*(1), 66. <https://doi.org/10.1186/0717-6287-47-66>

G

- Gallo, M., Coutinho, F. J., Vanner, R. J., Gayden, T., Mack, S. C., Murison, A., Remke, M., Li, R., Takayama, N., Desai, K., Lee, L., Lan, X., Park, N. I., Barsyte-Lovejoy, D., Smil, D., Sturm, D., Kushida, M. M., Head, R., Cusimano, M. D., ... Dirks, P. B. (2015). MLL5 Orchestrates a Cancer Self-Renewal State by Repressing the Histone Variant H3.3 and Globally Reorganizing Chromatin. *Cancer Cell*, 28(6), 715–729. <https://doi.org/10.1016/j.ccell.2015.10.005>
- Gannon, L. M., Cotter, M. B., & Quinn, C. M. (2014). The classification of invasive carcinoma of the breast. *Expert Review of Anticancer Therapy*, 13(8), 941–954. <https://doi.org/10.1586/14737140.2013.820577>
- Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., & Wang, Z. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proceedings of the National Academy of Sciences*, 105(18), 6656–6661. <https://doi.org/10.1073/pnas.0801802105>
- Gao, X., McDonald, J. T., Naidu, M., Hahnfeldt, P., & Hlatky, L. (2014). A Proposed Quantitative Index for Assessing the Potential Contribution of Reprogramming to Cancer Stem Cell Kinetics. *Stem Cells International*, 2014, 249309. <https://doi.org/10.1155/2014/249309>
- Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., Yao, C., Liu, X., Lu, Z., Xu, Z., Kang, L., Chen, J., Wang, H., Cai, T., & Gao, S. (2013). Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. *Cell Stem Cell*, 12(4), 453–469. <https://doi.org/10.1016/j.stem.2013.02.005>
- Garcia-Mayea, Y., Mir, C., Masson, F., Paciucci, R., & LLeonart, M. (2019). Insights into new mechanisms and models of cancer stem cell multidrug resistance. *Seminars in Cancer Biology*, 60, 166–180. <https://doi.org/10.1016/j.semcancer.2019.07.022>
- Gardiner-Garden, M., & Frommer, M. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Biology*, 196(2), 261–282. [https://doi.org/10.1016/0022-2836\(87\)90689-9](https://doi.org/10.1016/0022-2836(87)90689-9)
- Gaspar, J. M., & Hart, R. P. (2017). DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics*, 18(1), 528. <https://doi.org/10.1186/s12859-017-1909-0>
- Genereux, D. P., Johnson, W. C., Burden, A. F., Stöger, R., & Laird, C. D. (2008). Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Research*, 36(22), e150–e150. <https://doi.org/10.1093/nar/gkn691>
- Ghisolfi, L., Keates, A. C., Hu, X., Lee, D.-k., & Li, C. J. (2012). Ionizing Radiation Induces Stemness in Cancer Cells. *PLoS ONE*, 7(8), e43628. <https://doi.org/10.1371/journal.pone.0043628>
- Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., & Jacob, S. T. (2005). 5-Aza-Deoxycytidine Induces Selective Degradation of DNA Methyltransferase 1 by a Proteasomal Pathway That Requires the KEN Box, Bromo-Adjacent Homology Domain, and Nuclear Localization Signal. *Molecular and Cellular Biology*, 25(11), 4727–4741. <https://doi.org/10.1128/mcb.25.11.4727-4741.2005>
- Giatromanolaki, A., Sivridis, E., Fiska, A., & Koukourakis, M. I. (2011). The CD44+/CD24- phenotype relates to 'triple-negative' state and unfavorable prognosis in breast cancer patients. *Medical Oncology*, 28(3), 745–752. <https://doi.org/10.1007/s12032-010-9530-3>
- Gieringer, M., Gosepath, J., & Naim, R. (2011). Radiotherapy and wound healing: Principles, management and prospects (Review). *Oncology Reports*, 26(2), 299–307. <https://doi.org/10.3892/or.2011.1319>
- Gil, J., Marques-Pamies, M., Valassi, E., García-Martínez, A., Serra, G., Hostalot, C., Fajardo-Montañana, C., Carrato, C., Bernabeu, I., Marazuela, M., Rodríguez-Lloveras, H., Cámara, R., Salinas, I., Lamas, C., Biagetti, B., Simó-Servat, A., Webb, S. M., Picó, A., Jordà, M., & Puig-Domingo, M. (2022). Implications of Heterogeneity of Epithelial-Mesenchymal States in Acromegaly Therapeutic Pharmacologic Response. *Biomedicines*, 10(2), 460. <https://doi.org/10.3390/biomedicines10020460>
- Gimple, R. C., Bhargava, S., Dixit, D., & Rich, J. N. (2019). Glioblastoma stem cells: lessons from the tumor hierarchy in a lethal cancer. *Genes & Development*, 33(11-12), 591–609. <https://doi.org/10.1101/gad.324301.119>
- Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C. G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M. S., & Dontu, G. (2007). ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. *Cell Stem Cell*, 1(5), 555–567. <https://doi.org/10.1016/j.stem.2007.08.014>
- Ginno, P. A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A. H. F. M., Edenhofer, F., Smallwood, S. A., Krebs, A. R., & Schübeler, D. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nature Communications*, 11(1), 2680. <https://doi.org/10.1038/s41467-020-16354-x>

- Gitan, R. S., Shi, H., Chen, C.-M., Yan, P. S., & Huang, T. H.-M. (2002). Methylation-Specific Oligonucleotide Microarray: A New Potential for High-Throughput Methylation Analysis. *Genome Research*, 12(1), 158–164. <https://doi.org/10.1101/gr.202801>
- Giuli, M. V., Mancusi, A., Giuliani, E., Screpanti, I., & Checquolo, S. (2021). Notch signaling in female cancers: a multifaceted node to overcome drug resistance. *Cancer Drug Resistance*, 4(4), 805–836. <https://doi.org/10.20517/cdr.2021.53>
- Global Cancer Observatory. (n.d.). <https://gco.iarc.fr/>
- Godet, I., & Gilkes, D. M. (2017). BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. *Integrative Cancer Science and Therapeutics*, 4(1). <https://doi.org/10.15761/icst.1000228>
- GoldenGate® Assay for Methylation and BeadArray™ Technology. (2010). https://www.illumina.com/Documents/products/technotes/technote%5C_goldengate%5Ctextbackslash%5C_assay%5Ctextbackslash%5C_methylation.pdf
- Golebiewska, A., Brons, N. H., Bjerkvig, R., & Niclou, S. P. (2011). Critical Appraisal of the Side Population Assay in Stem Cell and Cancer Stem Cell Research. *Cell Stem Cell*, 8(2), 136–147. <https://doi.org/10.1016/j.stem.2011.01.007>
- Gonzalzo, M. L., Liang, G., Spruck, C. H., Zingg, J. M., Rideout, W. M., & Jones, P. A. (1997). Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer research*, 57(4), 594–9.
- Gonzalzo, M. L., & Jones, P. A. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Research*, 25(12), 2529–2531. <https://doi.org/10.1093/nar/25.12.2529>
- Gorodetska, I., Lukiyanchuk, V., Peitzsch, C., Kozeretska, I., & Dubrovskaya, A. (2019). BRCA1 and EZH2 cooperate in regulation of prostate cancer stem cell phenotype. *International Journal of Cancer*, 145(11), 2974–2985. <https://doi.org/10.1002/ijc.32323>
- Gorringe, K. L., & Fox, S. B. (2017). Ductal Carcinoma In Situ Biology, Biomarkers, and Diagnosis. *Frontiers in Oncology*, 7, 248. <https://doi.org/10.3389/fonc.2017.00248>
- Gotoh, O. (1983). Prediction of melting profiles and local Helix stability for sequenced DNA. *Advances in Biophysics*, 16, iii. [https://doi.org/10.1016/0065-227x\(83\)90007-2](https://doi.org/10.1016/0065-227x(83)90007-2)
- Granados, K., Poelchen, J., Novak, D., & Utikal, J. (2020). Cellular Reprogramming—A Model for Melanoma Cellular Plasticity. *International Journal of Molecular Sciences*, 21(21), 8274. <https://doi.org/10.3390/ijms21218274>
- Gravina, G. L., Festuccia, C., Marampon, F., Popov, V. M., Pestell, R. G., Zani, B. M., & Tombolini, V. (2010). Biological rationale for the use of DNA methyltransferase inhibitors as new strategy for modulation of tumor response to chemotherapy and radiation. *Molecular Cancer*, 9(1), 305. <https://doi.org/10.1186/1476-4598-9-305>
- Green, M. R., & Sambrook, J. (2018). Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harbor Protocols*, 2018(5), pdb.prot095133. <https://doi.org/10.1101/pdb.prot095133>
- Green, M. R., & Sambrook, J. (2019a). Nested Polymerase Chain Reaction (PCR). *Cold Spring Harbor Protocols*, 2019(2), pdb.prot095182. <https://doi.org/10.1101/pdb.prot095182>
- Green, M. R., & Sambrook, J. (2019b). Polymerase Chain Reaction. *Cold Spring Harbor Protocols*, 2019(6), pdb.top095109. <https://doi.org/10.1101/pdb.top095109>
- Green, M. R., & Sambrook, J. (2019c). Polymerase Chain Reaction (PCR) Amplification of GC-Rich Templates. *Cold Spring Harbor Protocols*, 2019(2), pdb.prot095141. <https://doi.org/10.1101/pdb.prot095141>
- Greer, J. M., & McCombe, P. A. (2012). The role of epigenetic mechanisms and processes in autoimmune disorders. *Biologics : Targets & Therapy*, 6, 307–327. <https://doi.org/10.2147/btt.s24067>
- Grunau, C., Clark, S. J., & Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*, 29(13), e65–e65. <https://doi.org/10.1093/nar/29.13.e65>
- Grunau, C., Schattevoy, R., Mache, N., & Rosenthal, A. (2000). MethTools—a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Research*, 28(5), 1053–1058. <https://doi.org/10.1093/nar/28.5.1053>
- Gruntman, E., Qi, Y., Slotkin, R. K., Roeder, T., Martienssen, R. A., & Sachidanandam, R. (2008). Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics*, 9(1), 371–371. <https://doi.org/10.1186/1471-2105-9-371>
- Gu, P., LeMenuet, D., Chung, A. C. -K., Mancini, M., Wheeler, D. A., & Cooney, A. J. (2005). Orphan Nuclear Receptor GCNF Is Required for the Repression of Pluripotency Genes during Retinoic Acid-Induced Embryonic Stem Cell Differentiation. *Molecular and Cellular Biology*, 25(19), 8507–8519. <https://doi.org/10.1128/mcb.25.19.8507-8519.2005>
- Gu, P., Xu, X., Menuet, D. L., Chung, A. C.-K., & Cooney, A. J. (2011). Differential Recruitment of Methyl CpG-Binding Domain Factors and DNA Methyltransferases by the Orphan Receptor Germ Cell Nuclear Factor Initiates the Repression and Silencing of Oct4. *Stem Cells (Dayton, Ohio)*, 29(7), 1041–1051. <https://doi.org/10.1002/stem.652>

- Guldberg, P., Worm, J., & Grønbaek, K. (2002). Profiling DNA methylation by melting analysis. *Methods*, 27(2), 121–127. [https://doi.org/10.1016/s1046-2023\(02\)00063-4](https://doi.org/10.1016/s1046-2023(02)00063-4)
- Guo, L., Cheng, X., Chen, H., Chen, C., Xie, S., Zhao, M., Liu, D., Deng, Q., Liu, Y., Wang, X., Chen, X., Wang, J., Yin, Z., Qi, S., Gao, J., Ma, Y., Guo, N., & Shi, M. (2019). Induction of breast cancer stem cells by M1 macrophages through Lin-28B-let-7-HMGA2 axis. *Cancer Letters*, 452, 213–225. <https://doi.org/10.1016/j.canlet.2019.03.032>
- Guo, W., Chung, W.-Y., Qian, M., Pellegrini, M., & Zhang, M. Q. (2014). Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic Acids Research*, 42(5), 3009–3016. <https://doi.org/10.1093/nar/gkt1306>
- Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., & Lander, E. S. (2011). Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. *Cell*, 146(4), 633–644. <https://doi.org/10.1016/j.cell.2011.07.026>
- Gupta, P. B., Pastushenko, I., Skibinski, A., Blanpain, C., & Kuperwasser, C. (2019). Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell*, 24(1), 65–78. <https://doi.org/10.1016/j.stem.2018.11.011>
- Gustafsson, M. V., Zheng, X., Pereira, T., Gradin, K., Jin, S., Lundkvist, J., Ruas, J. L., Poellinger, L., Lendahl, U., & Bondesson, M. (2005). Hypoxia Requires Notch Signaling to Maintain the Undifferentiated Cell State. *Developmental Cell*, 9(5), 617–628. <https://doi.org/10.1016/j.devcel.2005.09.010>

H

- Habib, S. J., Chen, B.-C., Tsai, F.-C., Anastassiadis, K., Meyer, T., Betzig, E., & Nusse, R. (2013). A Localized Wnt Signal Orients Asymmetric Stem Cell Division in Vitro. *Science*, 339(6126), 1445–1448. <https://doi.org/10.1126/science.1231077>
- Haff, L. A. (1994). Improved quantitative PCR using nested primers. *Genome Research*, 3(6), 332–337. <https://doi.org/10.1101/gr.3.6.332>
- Haider, T., Pandey, V., Banjare, N., Gupta, P. N., & Soni, V. (2020). Drug resistance in cancer: mechanisms and tackling strategies. *Pharmacological Reports*, 72(5), 1125–1151. <https://doi.org/10.1007/s43440-020-00138-7>
- Han, M., Liu, M., Wang, Y., Chen, X., Xu, J., Sun, Y., Zhao, L., Qu, H., Fan, Y., & Wu, C. (2012a). Antagonism of miR-21 Reverses Epithelial-Mesenchymal Transition and Cancer Stem Cell Phenotype through AKT/ERK1/2 Inactivation by Targeting PTEN. *PLoS ONE*, 7(6), e39520. <https://doi.org/10.1371/journal.pone.0039520>
- Han, M., Wang, Y., Liu, M., Bi, X., Bao, J., Zeng, N., Zhu, Z., Mo, Z., Wu, C., & Chen, X. (2012b). MiR-21 regulates epithelial-mesenchymal transition phenotype and hypoxia-inducible factor-1 α expression in third-sphere forming breast cancer stem cell-like cells. *Cancer Science*, 103(6), 1058–1064. <https://doi.org/10.1111/j.1349-7006.2012.02281.x>
- Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1), 31–46. <https://doi.org/10.1158/2159-8290.cd-21-1059>
- Hanahan, D., & Coussens, L. M. (2012). Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *Cancer Cell*, 21(3), 309–322. <https://doi.org/10.1016/j.ccr.2012.02.022>
- Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57–70. [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9)
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Harrison, J., Stirzaker, C., & Clark, S. J. (1998). Cytosines Adjacent to Methylated CpG Sites Can Be Partially Resistant to Conversion in Genomic Bisulfite Sequencing Leading to Methylation Artifacts. *Analytical Biochemistry*, 264(1), 129–132. <https://doi.org/10.1006/abio.1998.2833>
- Hartmann, L. C., Degnim, A. C., Santen, R. J., Dupont, W. D., & Ghosh, K. (2015). Atypical Hyperplasia of the Breast — Risk Assessment and Management Options. *The New England Journal of Medicine*, 372(1), 78–89. <https://doi.org/10.1056/nejmsr1407164>
- Hartung, T., Rhein, M., Kalmbach, N., Thau-Habermann, N., Naujock, M., Müschen, L., Frieling, H., Sterneckert, J., Hermann, A., Wegner, F., & Petri, S. (2021). Methylation and Expression of Mutant FUS in Motor Neurons Differentiated From Induced Pluripotent Stem Cells From ALS Patients. *Frontiers in Cell and Developmental Biology*, 9, 774751. <https://doi.org/10.3389/fcell.2021.774751>
- Hashimoto, O., Shimizu, K., Semba, S., Chiba, S., Ku, Y., Yokozaki, H., & Hori, Y. (2011). Hypoxia Induces Tumor Aggressiveness and the Expansion of CD133-Positive Cells in a Hypoxia-Inducible Factor-1 α -Dependent Manner in Pancreatic Cancer Cells. *Pathobiology*, 78(4), 181–192. <https://doi.org/10.1159/000325538>
- Hass, R., Ohe, J. v. d., & Ungefroren, H. (2020). Impact of the Tumor Microenvironment on Tumor Heterogeneity and Consequences for Cancer Cell Plasticity and Stemness. *Cancers*, 12(12), 3716. <https://doi.org/10.3390/cancers12123716>

- Hassiotou, F., & Geddes, D. (2013). Anatomy of the human mammary gland: Current status of knowledge. *Clinical Anatomy*, 26(1), 29–48. <https://doi.org/10.1002/ca.22165>
- Hatakeyama, D., Tierling, S., Kuzuhara, T., & Müller, U. (2013). Methods in Neuroethological Research, 151–171. https://doi.org/10.1007/978-4-431-54331-2_10
- Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., Tanaka, S., & Shiota, K. (2007). Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes to Cells*, 12(3), 387–396. <https://doi.org/10.1111/j.1365-2443.2007.01058.x>
- Haviland, J. S., Owen, J. R., Dewar, J. A., Agrawal, R. K., Barrett, J., Barrett-Lee, P. J., Dobbs, H. J., Hopwood, P., Lawton, P. A., Magee, B. J., Mills, J., Simmons, S., Sydenham, M. A., Venables, K., Bliss, J. M., Yarnold, J. R., & Group, o. b. o. t. S. T. (2013). The UK Standardisation of Breast Radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. *The Lancet Oncology*, 14(11), 1086–1094. [https://doi.org/10.1016/s1470-2045\(13\)70386-3](https://doi.org/10.1016/s1470-2045(13)70386-3)
- Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J. A., . . . Ren, B. (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. *Cell Stem Cell*, 6(5), 479–491. <https://doi.org/10.1016/j.stem.2010.03.018>
- Hawley, L. (2013). Principles of radiotherapy. *British Journal of Hospital Medicine*, 74(Sup11), C166–C169. <https://doi.org/10.12968/hmed.2013.74.sup11.c166>
- Hayatsu, H. (2008a). Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — A personal account. *Proceedings of the Japan Academy, Series B*, (8), 321–330. <https://doi.org/10.2183/pjab.84.321>
- Hayatsu, H. (2008b). The bisulfite genomic sequencing used in the analysis of epigenetic states, a technique in the emerging environmental genotoxicology research. *Mutation Research/Reviews in Mutation Research*, 659(1-2), 77–82. <https://doi.org/10.1016/j.mrrev.2008.04.003>
- Hayatsu, H., & Miller, R. C. (1972). The cleavage of DNA by the oxygen-dependent reaction of bisulfite. *Biochemical and Biophysical Research Communications*, 46(1), 120–124. [https://doi.org/10.1016/0006-291x\(72\)90638-9](https://doi.org/10.1016/0006-291x(72)90638-9)
- Hayatsu, H., Negishi, K., & Shiraishi, M. (2004). Accelerated bisulfite-deamination of cytosine in the genomic sequencing procedure for DNA methylation analysis. *Nucleic Acids Symposium Series*, 48(1), 261–262. <https://doi.org/10.1093/nass/48.1.261>
- Hayatsu, H., Wataya, Y., Kai, K., & Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, 9(14), 2858–2865. <https://doi.org/10.1021/bi00816a016>
- He, J., Xiu, L., Porre, P. D., Dass, R., & Thomas, X. (2014). Decitabine reduces transfusion dependence in older patients with acute myeloid leukemia: results from a post hoc analysis of a randomized phase III study. *Leukemia & Lymphoma*, 56(4), 1033–1042. <https://doi.org/10.3109/10428194.2014.951845>
- He, S., Sun, H., Lin, L., Zhang, Y., Chen, J., Liang, L., Li, Y., Zhang, M., Yang, X., Wang, X., Wang, F., Zhu, F., Chen, J., Pei, D., & Zheng, H. (2017). Passive DNA demethylation preferentially up-regulates pluripotency-related genes and facilitates the generation of induced pluripotent stem cells. *Journal of Biological Chemistry*, 292(45), 18542–18555. <https://doi.org/10.1074/jbc.m117.810457>
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., & Xu, G.-L. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, 333(6047), 1303–1307. <https://doi.org/10.1126/science.1210944>
- Hebestreit, K., Dugas, M., & Klein, H.-U. (2013). Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics*, 29(13), 1647–1653. <https://doi.org/10.1093/bioinformatics/btt263>
- Hecker, K. H., & Roux, K. H. (1996). High and Low Annealing Temperatures Increase Both Specificity and Yield in Touchdown and Stepdown PCR. *BioTechniques*, 20(3), 478–485. <https://doi.org/10.2144/19962003478>
- Heddleston, J. M., Li, Z., McLendon, R. E., Hjelmeland, A. B., & Rich, J. N. (2009). The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell Cycle*, 8(20), 3274–3284. <https://doi.org/10.4161/cc.8.20.9701>
- Helczynska, K., Kronblad, A., Jögi, A., Nilsson, E., Beckman, S., Landberg, G., & Pählman, S. (2003). Hypoxia promotes a dedifferentiated phenotype in ductal breast carcinoma in situ. *Cancer research*, 63(7), 1441–4.
- Helou, R. E., Wicinski, J., Guille, A., Adélaïde, J., Finetti, P., Bertucci, F., Chaffanet, M., Birnbaum, D., Charafe-Jauffret, E., & Ginestier, C. (2014). Brief Reports: A Distinct DNA Methylation Signature Defines Breast Cancer Stem Cells and Predicts Cancer Outcome. *STEM CELLS*, 32(11), 3031–3036. <https://doi.org/10.1002/stem.1792>
- Henke, W., Herdel, K., Jung, K., Schnorr, D., & Loening, S. A. (1997). Betaine Improves the PCR Amplification of GC-Rich DNA Sequences. *Nucleic Acids Research*, 25(19), 3957–3958. <https://doi.org/10.1093/nar/25.19.3957>

- Herman, J. G., Graff, J. R., Myöhänen, S., Nelkin, B. D., & Baylin, S. B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences*, *93*(18), 9821–9826. <https://doi.org/10.1073/pnas.93.18.9821>
- Hermann, P. C., Huber, S. L., Herrler, T., Aicher, A., Ellwart, J. W., Guba, M., Bruns, C. J., & Heeschen, C. (2007). Distinct Populations of Cancer Stem Cells Determine Tumor Growth and Metastatic Activity in Human Pancreatic Cancer. *Cell Stem Cell*, *1*(3), 313–323. <https://doi.org/10.1016/j.stem.2007.06.002>
- Hernandez, H. G., Tse, M. Y., Pang, S. C., Arboleda, H., & Forero, D. A. (2013). Optimizing methodologies for PCR-based DNA methylation analysis. *BioTechniques*, *55*(4), 181–197. <https://doi.org/10.2144/000114087>
- Hernández-Vargas, H., Ouzounova, M., Calvez-Kelm, F. L., Lambert, M.-P., McKay-Chopin, S., Tavtigian, S. V., Puisieux, A., Matar, C., & Herceg, Z. (2011). Methylome analysis reveals Jak-STAT pathway deregulation in putative breast cancer stem cells. *Epigenetics*, *6*(4), 428–439. <https://doi.org/10.4161/epi.6.4.14515>
- Herreros-Villanueva, M., Zhang, J.-S., Koenig, A., Abel, E. V., Smyrk, T. C., Bamlet, W. R., Narvajias, A. A.-M. d., Gomez, T. S., Simeone, D. M., Bujanda, L., & Billadeau, D. D. (2013). SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis*, *2*(8), e61–e61. <https://doi.org/10.1038/oncsis.2013.23>
- Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., Backlund, M. G., Yin, Y., Khramtsov, A. I., Bastein, R., Quackenbush, J., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, L., . . . Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology*, *8*(5), R76–R76. <https://doi.org/10.1186/gb-2007-8-5-r76>
- Heseding, H. M., Jahn, K., Eberlein, C. K., Wieting, J., Maier, H. B., Proskynitopoulos, P. J., Glahn, A., Bleich, S., Frieling, H., & Deest, M. (2022). Distinct promoter regions of the oxytocin receptor gene are hypomethylated in Prader-Willi syndrome and in Prader-Willi syndrome associated psychosis. *Translational Psychiatry*, *12*(1), 246. <https://doi.org/10.1038/s41398-022-02014-9>
- Hetzl, J., Foerster, A. M., Raidl, G., & Scheid, O. M. (2007). CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *The Plant Journal*, *51*(3), 526–536. <https://doi.org/10.1111/j.1365-313x.2007.03152.x>
- Hodges, E., Smith, A. D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M. Q., Ye, K., Bhattacharjee, A., Brizuela, L., McCombie, W. R., Wigler, M., Hannon, G. J., & Hicks, J. B. (2009). High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Research*, *19*(9), 1593–1605. <https://doi.org/10.1101/gr.095190.109>
- Hogle, W. P. (2006). The State of the Art in Radiation Therapy. *Seminars in Oncology Nursing*, *22*(4), 212–220. <https://doi.org/10.1016/j.soncn.2006.07.004>
- Holemon, H., Korshunova, Y., Ordway, J., Bedell, J., Citek, R., Lakey, N., Leon, J., Finney, M., McPherson, J., & Jeddleloh, J. (2007). MethylScreen: DNA methylation density monitoring using quantitative PCR. *BioTechniques*, *43*(5), 683–693. <https://doi.org/10.2144/000112597>
- Holland, J. D., Klaus, A., Garratt, A. N., & Birchmeier, W. (2013). Wnt signaling in stem and cancer stem cells. *Current Opinion in Cell Biology*, *25*(2), 254–264. <https://doi.org/10.1016/j.ceb.2013.01.004>
- Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). p53 Mutations in Human Cancers. *Science*, *253*(5015), 49–53. <https://doi.org/10.1126/science.1905840>
- Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, *45*(10), 1198–1206. <https://doi.org/10.1038/ng.2746>
- Hori, K., Sen, A., & Artavanis-Tsakonas, S. (2013). Notch signaling at a glance. *Journal of Cell Science*, *126*(10), 2135–2140. <https://doi.org/10.1242/jcs.127308>
- Hou, H., Sun, H., Lu, P., Ge, C., Zhang, L., Li, H., Zhao, F., Tian, H., Zhang, L., Chen, T., Yao, M., & Li, J. (2013). Tunicamycin Potentiates Cisplatin Anticancer Efficacy through the DPAGT1/Akt/ABCG2 Pathway in Mouse Xenograft Models of Human Hepatocellular Carcinoma. *Molecular Cancer Therapeutics*, *12*(12), 2874–2884. <https://doi.org/10.1158/1535-7163.mct-13-0201>
- Hoyt, M. A., Zhang, M., & Coffino, P. (2005). Probing the Ubiquitin/Proteasome System with Ornithine Decarboxylase, a Ubiquitin-Independent Substrate. *Methods in Enzymology*, *398*, 399–413. [https://doi.org/10.1016/s0076-6879\(05\)98033-6](https://doi.org/10.1016/s0076-6879(05)98033-6)
- Hu, J. L., Wang, W., Lan, X. L., Zeng, Z. C., Liang, Y. S., Yan, Y. R., Song, F. Y., Wang, F. F., Zhu, X. H., Liao, W. J., Liao, W. T., Ding, Y. Q., & Liang, L. (2019a). CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer. *Molecular Cancer*, *18*(1), 91. <https://doi.org/10.1186/s12943-019-1019-x>
- Hu, X., Ghisolfi, L., Keates, A. C., Zhang, J., Xiang, S., Lee, D.-k. K., & Li, C. J. (2012). Induction of cancer cell stemness by chemotherapy. *Cell cycle (Georgetown, Tex.)*, *11*(14), 2691–8. <https://doi.org/10.4161/cc.21021>

- Hu, Y., & Fu, L. (2012). Targeting cancer stem cells: a new therapy to cure cancer patients. *American journal of cancer research*, 2(3), 340–56.
- Hu, Y.-B., Yan, C., Mu, L., Mi, Y.-L., Zhao, H., Hu, H., Li, X.-L., Tao, D.-D., Wu, Y.-Q., Gong, J.-P., & Qin, J.-C. (2019b). Exosomal Wnt-induced dedifferentiation of colorectal cancer cells contributes to chemotherapy resistance. *Oncogene*, 38(11), 1951–1965. <https://doi.org/10.1038/s41388-018-0557-9>
- Hu, Y., Yan, C., Mu, L., Huang, K., Li, X., Tao, D., Wu, Y., & Qin, J. (2015). Fibroblast-Derived Exosomes Contribute to Chemoresistance through Priming Cancer Stem Cells in Colorectal Cancer. *PLoS ONE*, 10(5), e0125625. <https://doi.org/10.1371/journal.pone.0125625>
- Huang, H., Hou, J., Liu, K., Liu, Q., Shen, L., Liu, B., Lu, Q., Zhang, N., Che, L., Li, J., Jiang, S., Wang, B., Wen, Q., Hu, L., & Gao, J. (2021). RAB27A-dependent release of exosomes by liver cancer stem cells induces Nanog expression in their differentiated progenies and confersregorafenib resistance. *Journal of Gastroenterology and Hepatology*, 36(12), 3429–3437. <https://doi.org/10.1111/jgh.15619>
- Huang, J., Li, C., Wang, Y., Lv, H., Guo, Y., Dai, H., Wicha, M. S., Chang, A. E., & Li, Q. (2013a). Cytokine-induced killer (CIK) cells bound with anti-CD3/anti-CD133 bispecific antibodies target CD133high cancer stem cells in vitro and in vivo. *Clinical Immunology*, 149(1), 156–168. <https://doi.org/10.1016/j.clim.2013.07.006>
- Huang, P., Chen, A., He, W., Li, Z., Zhang, G., Liu, Z., Liu, G., Liu, X., He, S., Xiao, G., Huang, F., Stenvang, J., Br nner, N., Hong, A., & Wang, J. (2017). BMP-2 induces EMT and breast cancer stemness through Rb and CD44. *Cell Death Discovery*, 3(1), 17039. <https://doi.org/10.1038/cddiscovery.2017.39>
- Huang, S.-D., Yuan, Y., Tang, H., Liu, X.-H., Fu, C.-G., Cheng, H.-Z., Bi, J.-W., Yu, Y.-W., Gong, D.-J., Zhang, W., Chen, J., & Xu, Z.-Y. (2013b). Tumor Cells Positive and Negative for the Common Cancer Stem Cell Markers Are Capable of Initiating Tumor Growth and Generating Both Progenies. *PLoS ONE*, 8(1), e54579. <https://doi.org/10.1371/journal.pone.0054579>
- Huang, S. (2012). Tumor progression: Chance and necessity in Darwinian and Lamarckian somatic (mutationless) evolution. *Progress in Biophysics and Molecular Biology*, 110(1), 69–86. <https://doi.org/10.1016/j.pbiomolbio.2012.05.001>
- Huang, T. H.-M., Perry, M. R., & Laux, D. E. (1999). Methylation Profiling of CpG Islands in Human Breast Cancer Cells. *Human Molecular Genetics*, 8(3), 459–470. <https://doi.org/10.1093/hmg/8.3.459>
- Hyman, R. W., Jiang, H., Fukushima, M., & Davis, R. W. (2010). A direct comparison of the KBTM Basecaller and phred for identifying the bases from DNA sequencing using chain termination chemistry. *BMC Research Notes*, 3(1), 257–257. <https://doi.org/10.1186/1756-0500-3-257>
- Ibrahim, A. E. K., Thorne, N. P., Baird, K., Barbosa-Morais, N. L., Tavar e, S., Collins, V. P., Wyllie, A. H., Arends, M. J., & Brenton, J. D. (2006). MMAPSS: an optimized array-based method for assessing CpG island methylation. *Nucleic Acids Research*, 34(20), e136–e136. <https://doi.org/10.1093/nar/gkl551>
- Iliopoulos, D., Hirsch, H. A., & Struhl, K. (2009). An Epigenetic Switch Involving NF- B, Lin28, Let-7 MicroRNA, and IL6 Links Inflammation to Cell Transformation. *Cell*, 139(4), 693–706. <https://doi.org/10.1016/j.cell.2009.10.014>
- Iliopoulos, D., Hirsch, H. A., Wang, G., & Struhl, K. (2011). Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proceedings of the National Academy of Sciences*, 108(4), 1397–1402. <https://doi.org/10.1073/pnas.1018898108>
- Iliopoulos, D., Lindahl-Allen, M., Polytaichou, C., Hirsch, H. A., Tschlis, P. N., & Struhl, K. (2010). Loss of miR-200 Inhibition of Suz12 Leads to Polycomb-Mediated Repression Required for the Formation and Maintenance of Cancer Stem Cells. *Molecular Cell*, 39(5), 761–772. <https://doi.org/10.1016/j.molcel.2010.08.013>
- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genetics*, 6(9), e1001134. <https://doi.org/10.1371/journal.pgen.1001134>
- Illumina Methylation BeadChips Achieve Breadth of Coverage Using 2 Infinium Chemistries. (2015). https://www.illumina.com/documents/products/technotes/technote%5Ctextbackslash%5C_hm450%5Ctextbackslash%5C_data%5Ctextbackslash%5C_analysis%5Ctextbackslash%5C_optimization.pdf
- INCa, I. N. d. C. (2021). *Panorama des cancers en France -  dition 2021*. https://www.e-cancer.fr/pdf%5C_inca/preview/303372/4327939/file/Panorama%5C%20des%5C%20cancers%5C%20en%5C%20France%5C_2021.pdf
- Irizarry, R. A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S. A., Jeddloh, J. A., Wen, B., & Feinberg, A. P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Research*, 18(5), 780–790. <https://doi.org/10.1101/gr.7301508>
- Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S., & Feinberg, A. P. (2009). The human colon cancer methylome shows similar

- hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41(2), 178–186. <https://doi.org/10.1038/ng.298>
- Ismail, M. A., Samara, M., Sayab, A. A., Alsharshani, M., Yassin, M. A., Varadharaj, G., Vezzalini, M., Tomasello, L., Monne, M., Morsi, H., Qoronfleh, M. W., Zayed, H., Cook, R., Sorio, C., Modjtahedi, H., & Al-Dewik, N. I. (2020). Aberrant DNA methylation of PTPRG as one possible mechanism of its under-expression in CML patients in the State of Qatar. *Molecular Genetics & Genomic Medicine*, 8(10), e1319. <https://doi.org/10.1002/mgg3.1319>
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047), 1300–1303. <https://doi.org/10.1126/science.1210597>
- Izar, B., Rotow, J., Gainor, J., Clark, J., & Chabner, B. (2013). Pharmacokinetics, Clinical Indications, and Resistance Mechanisms in Molecular Targeted Therapies in Cancer. *Pharmacological Reviews*, 65(4), 1351–1395. <https://doi.org/10.1124/pr.113.007807>
- ## J
- Jabbari, N., Karimipour, M., Khaksar, M., Akbariazar, E., Heidarzadeh, M., Mojarad, B., Aftab, H., Rahbarghazi, R., & Rezaie, J. (2020). Tumor-derived extracellular vesicles: insights into bystander effects of exosomes after irradiation. *Lasers in Medical Science*, 35(3), 531–545. <https://doi.org/10.1007/s10103-019-02880-8>
- Jaeger, S., Igea, A., Arroyo, R., Alcalde, V., Canovas, B., Orozco, M., Nebreda, A. R., & Aloy, P. (2017). Quantification of Pathway Cross-talk Reveals Novel Synergistic Drug Combinations for Breast Cancer. *Cancer Research*, 77(2), 459–469. <https://doi.org/10.1158/0008-5472.can-16-0097>
- Jelinek, J., Liang, S., Estecio, M. R. H., He, R., Lu, Y., Ramagli, L. S., & Issa, J.-P. J. (2009). Digital Restriction Enzyme Analysis of Methylation (DREAM) by Next Generation Sequencing Yields High Resolution Maps of DNA Methylation. *Blood*, 114(22), 567–567. <https://doi.org/10.1182/blood.v114.22.567.567>
- Jella, K. K., Rani, S., O'Driscoll, L., McClean, B., Byrne, H. J., & Lyng, F. M. (2014). Exosomes Are Involved in Mediating Radiation Induced Bystander Signaling in Human Keratinocyte Cells. *Radiation Research*, 181(2), 138–145. <https://doi.org/10.1667/rr13337.1>
- Jelonek, K., Widlak, P., & Pietrowska, M. (2016). The Influence of Ionizing Radiation on Exosome Composition, Secretion and Intercellular Communication. *Protein and Peptide Letters*, 23(7), 656–663. <https://doi.org/10.2174/0929866523666160427105138>
- Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, 449(7159), 248–251. <https://doi.org/10.1038/nature06146>
- Jiang, C., & Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, 10(3), 161–172. <https://doi.org/10.1038/nrg2522>
- Jiang, M., Zhang, Y., Fei, J., Chang, X., Fan, W., Qian, X., Zhang, T., & Lu, D. (2010). Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. *Laboratory Investigation*, 90(2), 282–290. <https://doi.org/10.1038/labinvest.2009.132>
- Jiang, X., Yan, Y., Hu, M., Chen, X., Wang, Y., Dai, Y., Wu, D., Wang, Y., Zhuang, Z., & Xia, H. (2016). Increased level of H19 long noncoding RNA promotes invasion, angiogenesis, and stemness of glioblastoma cells. *Journal of Neurosurgery*, 2016(1), 129–136. <https://doi.org/10.3171/2014.12.jns1426.test>
- Jilkine, A. (2019). Mathematical Models of Stem Cell Differentiation and Dedifferentiation. *Current Stem Cell Reports*, 5(2), 66–72. <https://doi.org/10.1007/s40778-019-00156-z>
- Jin, W. (2020). Role of JAK/STAT3 Signaling in the Regulation of Metastasis, the Transition of Cancer Stem Cells, and Chemoresistance of Cancer by Epithelial–Mesenchymal Transition. *Cells*, 9(1), 217. <https://doi.org/10.3390/cells9010217>
- Jin, X., Jin, X., & Kim, H. (2017). Cancer stem cells and differentiation therapy. *Tumor Biology*, 39(10), 1010428317729933. <https://doi.org/10.1177/1010428317729933>
- Jing, N., Gao, W.-Q., & Fang, Y.-X. (2021). Regulation of Formation, Stemness and Therapeutic Resistance of Cancer Stem Cells. *Frontiers in Cell and Developmental Biology*, 9, 641498. <https://doi.org/10.3389/fcell.2021.641498>
- Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V., & Jordan, I. K. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget*, 3(4), 462–474. <https://doi.org/10.18632/oncotarget.497>
- Jolly, M. K., Mani, S. A., & Levine, H. (2018). Hybrid epithelial/mesenchymal phenotype(s): The 'fittest' for metastasis? *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1870(2), 151–157. <https://doi.org/10.1016/j.bbcan.2018.07.001>
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7), 484–492. <https://doi.org/10.1038/nrg3230>

- Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics*, *10*(11), 805–811. <https://doi.org/10.1038/nrg2651>
- Jones, P. A., & Takai, D. (2001). The Role of DNA Methylation in Mammalian Epigenetics. *Science*, *293*(5532), 1068–1070. <https://doi.org/10.1126/science.1063852>
- Joseph, C., Arshad, M., Kurozomi, S., Althobiti, M., Miligy, I. M., Al-izzi, S., Toss, M. S., Goh, F. Q., Johnston, S. J., Martin, S. G., Ellis, I. O., Mongan, N. P., Green, A. R., & Rakha, E. A. (2019). Overexpression of the cancer stem cell marker CD133 confers a poor prognosis in invasive breast cancer. *Breast Cancer Research and Treatment*, *174*(2), 387–399. <https://doi.org/10.1007/s10549-018-05085-9>
- Jühling, F., Kretzmer, H., Bernhart, S. H., Otto, C., Stadler, P. F., & Hoffmann, S. (2016). metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Research*, *26*(2), 256–262. <https://doi.org/10.1101/gr.196394.115>
- Junk, D. J., Cipriano, R., Bryson, B. L., Gilmore, H. L., & Jackson, M. W. (2013). Tumor Microenvironmental Signaling Elicits Epithelial-Mesenchymal Plasticity through Cooperation with Transforming Genetic Events. *Neoplasia*, *15*(9), 1100–1109. <https://doi.org/10.1593/neo.131114>
- Jüttermann, R., Li, E., & Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences*, *91*(25), 11797–11801. <https://doi.org/10.1073/pnas.91.25.11797>

K

- Kabakov, A. E., & Yakimova, A. O. (2021). Hypoxia-Induced Cancer Cell Responses Driving Radioresistance of Hypoxic Tumors: Approaches to Targeting and Radiosensitizing. *Cancers*, *13*(5), 1102. <https://doi.org/10.3390/cancers13051102>
- Kagara, N., Huynh, K. T., Kuo, C., Okano, H., Sim, M. S., Elashoff, D., Chong, K., Giuliano, A. E., & Hoon, D. S. (2012). Epigenetic Regulation of Cancer Stem Cell Genes in Triple-Negative Breast Cancer. *The American Journal of Pathology*, *181*(1), 257–267. <https://doi.org/10.1016/j.ajpath.2012.03.019>
- Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, *119*(6), 1420–1428. <https://doi.org/10.1172/jci39104>
- Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grandér, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., & Ekström, T. J. (2006). LUMA (LUminometric Methylation Assay)—A high throughput method to the analysis of genomic DNA methylation. *Experimental Cell Research*, *312*(11), 1989–1995. <https://doi.org/10.1016/j.yexcr.2006.03.006>
- Karsli-Ceppioglu, S., Dagdemir, A., Judes, G., Ngollo, M., Penault-Llorca, F., Pajon, A., Bignon, Y.-J., & Bernard-Gallon, D. (2014). Epigenetic mechanisms of breast cancer: an update of the current knowledge. *Epigenomics*, *6*(6), 651–664. <https://doi.org/10.2217/epi.14.59>
- Kashyap, V., Rezende, N. C., Scotland, K. B., Shaffer, S. M., Persson, J. L., Gudas, L. J., & Mongan, N. P. (2009). Regulation of Stem Cell Pluripotency and Differentiation Involves a Mutual Regulatory Circuit of the Nanog, OCT4, and SOX2 Pluripotency Transcription Factors With Polycomb Repressive Complexes and Stem Cell microRNAs. *Stem Cells and Development*, *18*(7), 1093–1108. <https://doi.org/10.1089/scd.2009.0113>
- Katoh, M. (2017). Canonical and non-canonical WNT signaling in cancer stem cells and their niches: Cellular heterogeneity, omics reprogramming, targeted therapy and tumor plasticity (Review). *International Journal of Oncology*, *51*(5), 1357–1369. <https://doi.org/10.3892/ijo.2017.4129>
- Katoh, M., & Katoh, M. (2007). WNT Signaling Pathway and Stem Cell Signaling Network. *Clinical Cancer Research*, *13*(14), 4042–4045. <https://doi.org/10.1158/1078-0432.ccr-06-2316>
- Kaup, S., Grandjean, V., Mukherjee, R., Kapoor, A., Keyes, E., Seymour, C. B., Mothersill, C. E., & Schofield, P. N. (2006). Radiation-induced genomic instability is associated with DNA methylation changes in cultured human keratinocytes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *597*(1-2), 87–97. <https://doi.org/10.1016/j.mrfmmm.2005.06.032>
- Kawaguchi-Ihara, N., Murohashi, I., Nara, N., & Tohda, S. (2008). Promotion of the self-renewal capacity of human acute leukemia cells by Wnt3A. *Anticancer research*, *28*(5A), 2701–4.
- Kawano, R., Ohta, K., & Lupo, G. (2017). Cadherin-7 enhances Sonic Hedgehog signalling by preventing Gli3 repressor formation during neural tube patterning. *Open Biology*, *7*(12), 170225. <https://doi.org/10.1098/rsob.170225>
- Keshamouni, V. G., Jagtap, P., Michailidis, G., Strahler, J. R., Kuick, R., Reka, A. K., Papoulias, P., Krishnapuram, R., Sri-rangam, A., Standiford, T. J., Andrews, P. C., & Omenn, G. S. (2009). Temporal Quantitative Proteomics by iTRAQ 2D-LC-MS/MS and Corresponding mRNA Expression Analysis Identify Post-Transcriptional Modulation of Actin-Cytoskeleton Regulators During TGF- β -Induced Epithelial-Mesenchymal Transition. *Journal of Proteome Research*, *8*(1), 35–47. <https://doi.org/10.1021/pr8006478>

- Kessel, K. E. v., Neste, L. V., Lurkin, I., Zwarthoff, E. C., & Criekinge, W. V. (2016). Evaluation of an Epigenetic Profile for the Detection of Bladder Cancer in Patients with Hematuria. *The Journal of Urology*, 195(3), 601–607. <https://doi.org/10.1016/j.juro.2015.08.085>
- Keymeulen, A. V., Rocha, A. S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., & Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature*, 479(7372), 189–193. <https://doi.org/10.1038/nature10573>
- Khodadadi, E., Fahmideh, L., Khodadadi, E., Dao, S., Yousefi, M., Taghizadeh, S., Asgharzadeh, M., Yousefi, B., & Kafil, H. S. (2021). Current Advances in DNA Methylation Analysis Methods. *BioMed Research International*, 2021, 8827516. <https://doi.org/10.1155/2021/8827516>
- Khong, T., Sharkey, J., & Spencer, A. (2008). The effect of azacitidine on interleukin-6 signaling and nuclear factor- κ B activation and its in vitro and in vivo activity against multiple myeloma. *Haematologica*, 93(6), 860–869. <https://doi.org/10.3324/haematol.12261>
- Khulan, B., Thompson, R. F., Ye, K., Fazzari, M. J., Suzuki, M., Stasiak, E., Figueroa, M. E., Glass, J. L., Chen, Q., Montagna, C., Hatchwell, E., Selzer, R. R., Richmond, T. A., Green, R. D., Melnick, A., & Grealley, J. M. (2006). Comparative isoschizomer profiling of cytosine methylation: The HELP assay. *Genome Research*, 16(8), 1046–1055. <https://doi.org/10.1101/gr.5273806>
- Kim, B., Na, Y., Kim, J., Jeong, Y. A., Park, S., Jo, M., Jeong, S., Kang, S., Oh, S., & Lee, D.-H. (2019a). RUNX3 suppresses metastasis and stemness by inhibiting Hedgehog signaling in colorectal cancer. *Cell Death & Differentiation*, 1–19. <https://doi.org/10.1038/s41418-019-0379-5>
- Kim, D.-Y., Park, E., Heo, C. Y., Jin, U. S., Kim, E. K., Han, W., Shin, K. H., & Kim, I. A. (2021). Hypofractionated versus conventional fractionated radiotherapy for breast cancer in patients with reconstructed breast: Toxicity analysis. *The Breast*, 55, 37–44. <https://doi.org/10.1016/j.breast.2020.11.020>
- Kim, H., Kim, J., Chie, E., DaYoung, P., Kim, I., & Kim, I. (2012). DNMT (DNA methyltransferase) inhibitors radiosensitize human cancer cells by suppressing DNA repair activity. *Radiation Oncology*, 7(1), 39. <https://doi.org/10.1186/1748-717x-7-39>
- Kim, H. J., Kim, M.-J., Ahn, S. H., Son, B. H., Kim, S. B., Ahn, J. H., Noh, W. C., & Gong, G. (2011). Different prognostic significance of CD24 and CD44 expression in breast cancer according to hormone receptor status. *The Breast*, 20(1), 78–85. <https://doi.org/10.1016/j.breast.2010.08.001>
- Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M., Ji, H., Ehrlich, L., Yabuuchi, A., Takeuchi, A., Cunliffe, K., Hongguang, H., McKinney-Freeman, S., Naveiras, O., Yoon, T., Irizarry, R., Jung, N., . . . Daley, G. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature*, 467(7313), 285–290. <https://doi.org/10.1038/nature09342>
- Kim, M., & Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. *Experimental & molecular medicine*, 49(4), e322. <https://doi.org/10.1038/emm.2017.10>
- Kim, S. K., Kim, H., Lee, D.-h., Kim, T.-s., Kim, T., Chung, C., Koh, G. Y., Kim, H., & Lim, D.-S. (2013a). Reversing the Intractable Nature of Pancreatic Cancer by Selectively Targeting ALDH-High, Therapy-Resistant Cancer Cells. *PLoS ONE*, 8(10), e78130. <https://doi.org/10.1371/journal.pone.0078130>
- Kim, S.-Y., Kang, J., Song, X., Kim, B., Yoo, Y., Kwon, Y., & Lee, Y. J. (2013b). Role of the IL-6-JAK1-STAT3-Oct-4 pathway in the conversion of non-stem cancer cells into cancer stem-like cells. *Cellular Signalling*, 25(4), 961–969. <https://doi.org/10.1016/j.cellsig.2013.01.007>
- Kim, W., Lee, S., Seo, D., Kim, D., Kim, K., Kim, E., Kang, J., Seong, K. M., Youn, H., & Youn, B. (2019b). Cellular Stress Responses in Radiotherapy. *Cells*, 8(9), 1105. <https://doi.org/10.3390/cells8091105>
- Kischel, P., Bellahcene, A., Deux, B., Lamour, V., Dobson, R., Pauw, E. D., Clezardin, P., & Castronovo, V. (2012). Overexpression of CD9 in human breast cancer cells promotes the development of bone metastases. *Anticancer research*, 32(12), 5211–20.
- Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.-A., Mitterecker, A., Bodenhofer, U., & Hochreiter, S. (2012). cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Research*, 40(9), e69–e69. <https://doi.org/10.1093/nar/gks003>
- Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, 31(2), 89–97. <https://doi.org/10.1016/j.tibs.2005.12.008>
- Knudson, A. G. (2001). Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, 1(2), 157–162. <https://doi.org/10.1038/35101031>
- Koch, A., Joosten, S. C., Feng, Z., Ruijter, T. C. d., Draht, M. X., Melotte, V., Smits, K. M., Veeck, J., Herman, J. G., Neste, L. V., Criekinge, W. V., Meyer, T. D., & Engeland, M. v. (2018). Analysis of DNA methylation in cancer: location revisited. *Nature Reviews Clinical Oncology*, 15(7), 459–466. <https://doi.org/10.1038/s41571-018-0004-4>

- Kong, D., Ahmad, A., Bao, B., Li, Y., Banerjee, S., & Sarkar, F. H. (2012). Histone Deacetylase Inhibitors Induce Epithelial-to-Mesenchymal Transition in Prostate Cancer Cells. *PLoS ONE*, 7(9), e45045. <https://doi.org/10.1371/journal.pone.0045045>
- Kong, L., Guo, S., Liu, C., Zhao, Y., Feng, C., Liu, Y., Wang, T., & Li, C. (2016). Overexpression of SDF-1 activates the NF- κ B pathway to induce epithelial to mesenchymal transition and cancer stem cell-like phenotypes of breast cancer cells. *International Journal of Oncology*, 48(3), 1085–1094. <https://doi.org/10.3892/ijo.2016.3343>
- Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3(9), 1452–1456. <https://doi.org/10.1038/nprot.2008.133>
- Korkaya, H., Kim, G.-i., Davis, A., Malik, F., Henry, N. L., Ithimakin, S., Quraishi, A. A., Tawakkol, N., D'Angelo, R., Paulson, A. K., Chung, S., Luther, T., Paholak, H. J., Liu, S., Hassan, K. A., Zen, Q., Clouthier, S. G., & Wicha, M. S. (2012). Activation of an IL6 Inflammatory Loop Mediates Trastuzumab Resistance in HER2+ Breast Cancer by Expanding the Cancer Stem Cell Population. *Molecular Cell*, 47(4), 570–584. <https://doi.org/10.1016/j.molcel.2012.06.014>
- Korkaya, H., Paulson, A., Charafe-Jauffret, E., Ginestier, C., Brown, M., Dutcher, J., Clouthier, S. G., & Wicha, M. S. (2009). Regulation of Mammary Stem/Progenitor Cells by PTEN/Akt/ β -Catenin Signaling. *PLoS Biology*, 7(6), e1000121. <https://doi.org/10.1371/journal.pbio.1000121>
- Korpal, M., Lee, E. S., Hu, G., & Kang, Y. (2008). The miR-200 Family Inhibits Epithelial-Mesenchymal Transition and Cancer Cell Migration by Direct Targeting of E-cadherin Transcriptional Repressors ZEB1 and ZEB2 *. *Journal of Biological Chemistry*, 283(22), 14910–14914. <https://doi.org/10.1074/jbc.c800074200>
- Krebs, A. M., Mitschke, J., Losada, M. L., Schmalhofer, O., Boerries, M., Busch, H., Boettcher, M., Mougiakakos, D., Reichardt, W., Bronsert, P., Brunton, V. G., Pilarsky, C., Winkler, T. H., Brabletz, S., Stemmler, M. P., & Brabletz, T. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature Cell Biology*, 19(5), 518–529. <https://doi.org/10.1038/ncb3513>
- Krinner, S., Heitzer, A. P., Diermeier, S. D., Obermeier, I., Längst, G., & Wagner, R. (2014). CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Research*, 42(6), 3551–3564. <https://doi.org/10.1093/nar/gkt1358>
- Kristensen, L. S., Mikeska, T., Krypuy, M., & Dobrovic, A. (2008). Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Research*, 36(7), e42–e42. <https://doi.org/10.1093/nar/gkn113>
- Kröger, C., Afeyan, A., Mraz, J., Eaton, E. N., Reinhardt, F., Khodor, Y. L., Thiru, P., Bierie, B., Ye, X., Burge, C. B., & Weinberg, R. A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 116(15), 7353–7362. <https://doi.org/10.1073/pnas.1812876116>
- Kuhmann, C., Weichenhan, D., Rehli, M., Plass, C., Schmezer, P., & Popanda, O. (2011). DNA methylation changes in cells regrowing after fractioned ionizing radiation. *Radiotherapy and Oncology*, 101(1), 116–121. <https://doi.org/10.1016/j.radonc.2011.05.048>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(Web Server issue), W90–W97. <https://doi.org/10.1093/nar/gkw377>
- Kulis, M., & Esteller, M. (2010). 2 DNA Methylation and Cancer. *Advances in Genetics*, 70, 27–56. <https://doi.org/10.1016/b978-0-12-380866-0.60002-2>
- Kulkarni, M., Tan, T. Z., Sulaiman, N. B. S., Lamar, J. M., Bansal, P., Cui, J., Qiao, Y., & Ito, Y. (2018). RUNX1 and RUNX3 protect against YAP-mediated EMT, stem-ness and shorter survival outcomes in breast cancer. *Oncotarget*, 9(18), 14175–14192. <https://doi.org/10.18632/oncotarget.24419>
- Kumaki, Y., Oda, M., & Okano, M. (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Research*, 36(Web Server issue), W170–W175. <https://doi.org/10.1093/nar/gkn294>
- Kumar, N., Patni, P., Agarwal, A., Khan, M., & Parashar, N. (2015). Prevalence of molecular subtypes of invasive breast cancer: A retrospective study. *Medical Journal Armed Forces India*, 71(3), 254–258. <https://doi.org/10.1016/j.mjafi.2015.04.006>
- Kundu, P., Torres, E. R. S., Stagaman, K., Kasschau, K., Okhovat, M., Holden, S., Ward, S., Nevenon, K. A., Davis, B. A., Saito, T., Saido, T. C., Carbone, L., Sharpton, T. J., & Raber, J. (2021). Integrated analysis of behavioral, epigenetic, and gut microbiome analyses in AppNL-G-F, AppNL-F, and wild type mice. *Scientific Reports*, 11(1), 4678. <https://doi.org/10.1038/s41598-021-83851-4>
- Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R., & Ehrlich, M. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Research*, 8(20), 4763–4776. <https://doi.org/10.1093/nar/8.20.4763>

- Kurdyukov, S., & Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology*, 5(1), 3. <https://doi.org/10.3390/biology5010003>
- Kwok, P.-Y., Carlson, C., Yager, T. D., Ankener, W., & Nickerson, D. A. (1994). Comparative Analysis of Human DNA Variations by Fluorescence-Based Sequencing of PCR Products. *Genomics*, 23(1), 138–144. <https://doi.org/10.1006/geno.1994.1469>
- Kyjacova, L., Hubackova, S., Krejcikova, K., Strauss, R., Hanzlikova, H., Dzijak, R., Imrichova, T., Simova, J., Reinis, M., Bartek, J., & Hodny, Z. (2015). Radiotherapy-induced plasticity of prostate cancer mobilizes stem-like non-adherent, Erk signaling-dependent cells. *Cell Death & Differentiation*, 22(6), 898–911. <https://doi.org/10.1038/cdd.2014.97>

L

- Lagadec, C., Vlashi, E., Alhiyari, Y., Phillips, T. M., Dratver, M. B., & Pajonk, F. (2013). Radiation-Induced Notch Signaling in Breast Cancer Stem Cells. *International Journal of Radiation Oncology*Biophysics*, 87(3), 609–618. <https://doi.org/10.1016/j.ijrobp.2013.06.2064>
- Lagadec, C., Vlashi, E., Donna, L. D., Dekmezian, C., & Pajonk, F. (2012). Radiation-Induced Reprogramming of Breast Cancer Cells. *STEM CELLS*, 30(5), 833–844. <https://doi.org/10.1002/stem.1058>
- Lagadec, C., Vlashi, E., Donna, L. D., Meng, Y., Dekmezian, C., Kim, K., & Pajonk, F. (2010). Survival and self-renewing capacity of breast cancer initiating cells during fractionated radiation treatment. *Breast Cancer Research : BCR*, 12(1), R13–R13. <https://doi.org/10.1186/bcr2479>
- Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics*, 11(3), 191–203. <https://doi.org/10.1038/nrg2732>
- Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G., & Cedar, H. (2007). Role of DNA Methylation in Stable Gene Repression. *Journal of Biological Chemistry*, 282(16), 12194–12200. <https://doi.org/10.1074/jbc.m607838200>
- Landsberg, J., Kohlmeyer, J., Renn, M., Bald, T., Rogava, M., Cron, M., Fatho, M., Lennerz, V., Wölfel, T., Hölzel, M., & Tüting, T. (2012). Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*, 490(7420), 412–416. <https://doi.org/10.1038/nature11538>
- Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., & Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464), 645–648. <https://doi.org/10.1038/367645a0>
- Larson, S. M., Carrasquillo, J. A., Cheung, N.-K. V., & Press, O. W. (2015). Radioimmunotherapy of human tumours. *Nature Reviews Cancer*, 15(6), 347–360. <https://doi.org/10.1038/nrc3925>
- Lau, Q. C., Raja, E., Salto-Tellez, M., Liu, Q., Ito, K., Inoue, M., Putti, T. C., Loh, M., Ko, T. K., Huang, C., Bhalla, K. N., Zhu, T., Ito, Y., & Sukumar, S. (2006). RUNX3 Is Frequently Inactivated by Dual Mechanisms of Protein Mislocalization and Promoter Hypermethylation in Breast Cancer. *Cancer Research*, 66(13), 6512–6520. <https://doi.org/10.1158/0008-5472.can-06-0369>
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Sung, K. W. K., Rigoutsos, I., Loring, J., & Wei, C.-L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, 20(3), 320–331. <https://doi.org/10.1101/gr.101907.109>
- Lecerf, C., Peperstraete, E., Bourhis, X. L., & Adriaenssens, E. (2020). Propagation and Maintenance of Cancer Stem Cells: A Major Influence of the Long Non-Coding RNA H19. *Cells*, 9(12), 2613. <https://doi.org/10.3390/cells9122613>
- Lee, E.-J., Rath, P., Liu, J., Ryu, D., Pei, L., Noonpalle, S. K., Shull, A. Y., Feng, Q., Litofsky, S. N., Miller, D. C., Anthony, D. C., Kirk, M. D., Laterra, J., Deng, L., Xin, H.-B., Wang, X., Choi, J.-H., & Shi, H. (2015). Identification of Global DNA Methylation Signatures in Glioblastoma-Derived Cancer Stem Cells. *Journal of Genetics and Genomics*, 42(7), 355–371. <https://doi.org/10.1016/j.jgg.2015.06.003>
- Lee, E. Y., & Muller, W. J. (2010). Oncogenes and Tumor Suppressor Genes. *Cold Spring Harbor Perspectives in Biology*, 2(10), a003236. <https://doi.org/10.1101/cshperspect.a003236>
- Lee, G., Auffinger, B., Guo, D., Hasan, T., Deheeger, M., Tobias, A. L., Kim, J. Y., Atashi, F., Zhang, L., Lesniak, M. S., James, C. D., & Ahmed, A. U. (2016a). Dedifferentiation of Glioma Cells to Glioma Stem-like Cells By Therapeutic Stress-induced HIF Signaling in the Recurrent GBM Model. *Molecular Cancer Therapeutics*, 15(12), 3064–3076. <https://doi.org/10.1158/1535-7163.mct-15-0675>
- Lee, G., Hall, R. R., & Ahmed, A. U. (2016b). Cancer Stem Cells: Cellular Plasticity, Niche, and its Clinical Relevance. *Journal of stem cell research & therapy*, 6(10), 363. <https://doi.org/10.4172/2157-7633.1000363>
- Lee, H. E., Kim, J. H., Kim, Y. J., Choi, S. Y., Kim, S.-W., Kang, E., Chung, I. Y., Kim, I. A., Kim, E. J., Choi, Y., Ryu, H. S., & Park, S. Y. (2011). An increase in cancer stem cell population after primary systemic therapy is a poor prognostic factor in breast cancer. *British Journal of Cancer*, 104(11), 1730–1738. <https://doi.org/10.1038/bjc.2011.159>

- Lee, J. H., & Nan, A. (2012). Combination Drug Delivery Approaches in Metastatic Breast Cancer. *Journal of Drug Delivery*, 2012, 915375. <https://doi.org/10.1155/2012/915375>
- Lee, K.-S., Lee, Y.-S., Lee, J.-M., Ito, K., Cinghu, S., Kim, J.-H., Jang, J.-W., Li, Y.-H., Goh, Y.-M., Chi, X.-Z., Wee, H., Lee, H.-W., Hosoya, A., Chung, J.-H., Jang, J.-J., Kundu, J. K., Surh, Y.-J., Kim, W.-J., Ito, Y., ... Bae, S.-C. (2010). Runx3 is required for the differentiation of lung epithelial cells and suppression of lung cancer. *Oncogene*, 29(23), 3349–3361. <https://doi.org/10.1038/onc.2010.79>
- Lengauer, C., Kinzler, K. W., & Vogelstein, B. (1998). Genetic instabilities in human cancers. *Nature*, 396(6712), 643–649. <https://doi.org/10.1038/25292>
- Leon-Ferre, R. A., Hieken, T. J., & Boughey, J. C. (2021). The Landmark Series: Neoadjuvant Chemotherapy for Triple-Negative and HER2-Positive Breast Cancer. *Annals of Surgical Oncology*, 28(4), 2111–2119. <https://doi.org/10.1245/s10434-020-09480-9>
- Leontiou, C. A., Hadjidaniel, M. D., Mina, P., Antoniou, P., Ioannides, M., & Patsalis, P. C. (2015). Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers that Have the Potential to Be Used in Non-Invasive Prenatal Testing. *PLoS ONE*, 10(8), e0135058. <https://doi.org/10.1371/journal.pone.0135058>
- Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3), 203–21. <https://doi.org/10.1093/oxfordjournals.molbev.a040442>
- Lewin, J., Schmitt, A. O., Adorján, P., Hildmann, T., & Piepenbrock, C. (2004). Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics*, 20(17), 3005–3012. <https://doi.org/10.1093/bioinformatics/bth346>
- Lewis, J. S., & Jordan, V. C. (2005). Selective estrogen receptor modulators (SERMs): Mechanisms of anticarcinogenesis and drug resistance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 591(1-2), 247–263. <https://doi.org/10.1016/j.mrfmmm.2005.02.028>
- Li, C., Heidt, D. G., Dalerba, P., Burant, C. F., Zhang, L., Adsay, V., Wicha, M., Clarke, M. F., & Simeone, D. M. (2007a). Identification of Pancreatic Cancer Stem Cells. *Cancer Research*, 67(3), 1030–1037. <https://doi.org/10.1158/0008-5472.can-06-2030>
- Li, C., Fan, Y., Li, G., Xu, X., Duan, J., Li, R., Kang, X., Ma, X., Chen, X., Ke, Y., Yan, J., Lian, Y., Liu, P., Zhao, Y., Zhao, H., Chen, Y., Yu, Y., & Liu, J. (2018a). DNA methylation reprogramming of functional elements during mammalian embryonic development. *Cell Discovery*, 4(1), 41. <https://doi.org/10.1038/s41421-018-0039-9>
- Li, F., Zhou, K., Gao, L., Zhang, B., Li, W., Yan, W., Song, X., Yu, H., Wang, S., Yu, N., & Jiang, Q. (2016). Radiation induces the generation of cancer stem cells: A novel mechanism for cancer radioresistance. *Oncology Letters*, 12(5), 3059–3065. <https://doi.org/10.3892/ol.2016.5124>
- Li, G., Wang, D., Ma, W., An, K., Liu, Z., Wang, X., Yang, C., Du, F., Han, X., Chang, S., Yu, H., Zhang, Z., Zhao, Z., Zhang, Y., Wang, J., & Sun, Y. (2018b). Transcriptomic and epigenetic analysis of breast cancer stem cells. *Epigenomics*, 10(6), 765–783. <https://doi.org/10.2217/epi-2018-0008>
- Li, J., Lu, J., Ye, Z., Han, X., Zheng, X., Hou, H., Chen, W., Li, X., & Zhao, L. (2017a). 20(S)-Rg3 blocked epithelial-mesenchymal transition through DNMT3A/miR-145/FSCN1 in ovarian cancer. *Oncotarget*, 8(32), 53375–53386. <https://doi.org/10.18632/oncotarget.18482>
- Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., & Xu, G.-L. (2007b). Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of Oct4 and Nanog. *Molecular and Cellular Biology*, 27(24), 8748–8759. <https://doi.org/10.1128/mcb.01380-07>
- Li, L.-C. (2007). PCR Primer Design. *Methods in Molecular Biology™*, 402, 370–383. https://doi.org/10.1007/978-1-59745-528-2_19
- Li, L.-C., & Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, 18(11), 1427–1431. <https://doi.org/10.1093/bioinformatics/18.11.1427>
- Li, M., & Belmonte, J. C. I. (2017). Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics*, 18(3), 180–191. <https://doi.org/10.1038/nrg.2016.156>
- Li, M., & Belmonte, J. C. I. (2018). Deconstructing the pluripotency gene regulatory network. *Nature Cell Biology*, 20(4), 382–392. <https://doi.org/10.1038/s41556-018-0067-6>
- Li, S., Zhang, J., Huang, S., & He, X. (2017b). Genome-wide analysis reveals that exon methylation facilitates its selective usage in the human transcriptome. *Briefings in Bioinformatics*, 19(5), 754–764. <https://doi.org/10.1093/bib/bbx019>
- Li, S., & Tollefsbol, T. O. (2020). DNA methylation methods: global DNA methylation and methylomic analyses. *Methods*, 187, 28–43. <https://doi.org/10.1016/j.ymeth.2020.10.002>
- Li, Y., Wang, Z., Ajani, J. A., & Song, S. (2021). Drug resistance and Cancer stem cells. *Cell Communication and Signaling : CCS*, 19(1), 19. <https://doi.org/10.1186/s12964-020-00627-5>

- Li, Y., & Tollefsbol, T. O. (2011). Epigenetics Protocols. *Methods in Molecular Biology*, 791, 11–21. https://doi.org/10.1007/978-1-61779-316-5_2
- Liang, Y., Yang, L., & Xie, J. (2021). The Role of the Hedgehog Pathway in Chemoresistance of Gastrointestinal Cancers. *Cells*, 10(8), 2030. <https://doi.org/10.3390/cells10082030>
- Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., & Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genetics*, 43(11), 1091–1097. <https://doi.org/10.1038/ng.946>
- Lim, Y.-Y., Wright, J. A., Attema, J. L., Gregory, P. A., Bert, A. G., Smith, E., Thomas, D., Lopez, A. F., Drew, P. A., Khew-Goodall, Y., & Goodall, G. J. (2013). Epigenetic modulation of the miR-200 family is associated with transition to a breast cancer stem-cell-like state. *Journal of Cell Science*, 126(10), 2256–2266. <https://doi.org/10.1242/jcs.122275>
- Lin, P.-C., Hsieh, H.-Y., Chu, P.-C., & Chen, C. S. (2018). Therapeutic Opportunities of Targeting Histone Deacetylase Isoforms to Eradicate Cancer Stem Cells. *International Journal of Molecular Sciences*, 19(7), 1939. <https://doi.org/10.3390/ijms19071939>
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., & Ecker, J. R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell*, 133(3), 523–536. <https://doi.org/10.1016/j.cell.2008.03.029>
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315–322. <https://doi.org/10.1038/nature08514>
- Liu, A., Yu, X., & Liu, S. (2013a). Pluripotency transcription factors and cancer stem cells: small genes make a big difference. *Chinese Journal of Cancer*, 32(9), 483–487. <https://doi.org/10.5732/cjc.012.10282>
- Liu, B., Du, R., Zhou, L., Xu, J., Chen, S., Chen, J., Yang, X., Liu, D.-x., Shao, Z.-m., Zhang, L., Yu, Z., Xie, N., Guan, J.-l., & Liu, S. (2018a). miR-200c/141 Regulates Breast Cancer Stem Cell Heterogeneity via Targeting HIPK1/β-Catenin Axis. *Theranostics*, 8(21), 5801–5813. <https://doi.org/10.7150/thno.29380>
- Liu, C.-C., Lin, J.-H., Hsu, T.-W., Su, K., Li, A. F.-Y., Hsu, H.-S., & Hung, S.-C. (2014). IL-6 enriched lung cancer stem-like cell population by inhibition of cell cycle regulators via DNMT1 upregulation. *International Journal of Cancer*, 136(3), 547–559. <https://doi.org/10.1002/ijc.29033>
- Liu, C., Liu, L., Chen, X., Cheng, J., Zhang, H., Shen, J., Shan, J., Xu, Y., Yang, Z., Lai, M., & Qian, C. (2016). Sox9 regulates self-renewal and tumorigenicity by promoting symmetrical cell division of cancer stem cells in hepatocellular carcinoma. *Hepatology*, 64(1), 117–129. <https://doi.org/10.1002/hep.28509>
- Liu, G., Yuan, X., Zeng, Z., Tunici, P., Ng, H., Abdulkadir, I. R., Lu, L., Irvin, D., Black, K. L., & Yu, J. S. (2006a). Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Molecular Cancer*, 5(1), 67–67. <https://doi.org/10.1186/1476-4598-5-67>
- Liu, G., Ouyang, X., Gong, L., Yao, L., Liu, S., Li, J., Zhang, Q., & Xiao, Y. (2021a). E2F3 promotes liver cancer progression under the regulation of circ-PRKAR1B. *Molecular Therapy - Nucleic Acids*, 26, 104–113. <https://doi.org/10.1016/j.omtn.2021.07.002>
- Liu, H., Chen, C., Ma, D., Li, Y., Yin, Q., Li, Q., & Xiang, C. (2020a). Inhibition of PIM1 attenuates the stem cell-like traits of breast cancer cells by promoting RUNX3 nuclear retention. *Journal of Cellular and Molecular Medicine*, 24(11), 6308–6323. <https://doi.org/10.1111/jcmm.15272>
- Liu, H., Yan, Z., Yin, Q., Cao, K., Wei, Y., Rodriguez-Canales, J., Ma, D., & Wu, Y. (2018b). RUNX3 Epigenetic Inactivation Is Associated With Estrogen Receptor Positive Breast Cancer. *Journal of Histochemistry & Cytochemistry*, 66(10), 709–721. <https://doi.org/10.1369/0022155418797315>
- Liu, K., Lin, B., Zhao, M., Yang, X., Chen, M., Gao, A., Liu, F., Que, J., & Lan, X. (2013b). The multiple roles for Sox2 in stem cell maintenance and tumorigenesis. *Cellular Signalling*, 25(5), 1264–1271. <https://doi.org/10.1016/j.cellsig.2013.02.013>
- Liu, L., Zhang, Z., Zhou, L., Hu, L., Yin, C., Qing, D., Huang, S., Cai, X., & Chen, Y. (2020b). Cancer associated fibroblasts-derived exosomes contribute to radioresistance through promoting colorectal cancer stem cells phenotype. *Experimental Cell Research*, 391(2), 111956. <https://doi.org/10.1016/j.yexcr.2020.111956>
- Liu, S., Cheng, K., Zhang, H., Kong, R., Wang, S., Mao, C., & Liu, S. (2020c). Methylation Status of the Nanog Promoter Determines the Switch between Cancer Cells and Cancer Stem Cells. *Advanced Science*, 7(5), 1903035. <https://doi.org/10.1002/advs.201903035>
- Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S. P., Landis, M. D., Hong, S., Adams, A., D'Angelo, R., Ginestier, C., Charafe-Jauffret, E., Clouthier, S. G., Birnbaum, D., Wong, S. T., Zhan, M., . . . Wicha, M. S. (2013c). Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts. *Stem Cell Reports*, 2(1), 78–91. <https://doi.org/10.1016/j.stemcr.2013.11.009>

- Liu, S., Dontu, G., Mantle, I. D., Patel, S., Ahn, N.-s., Jackson, K. W., Suri, P., & Wicha, M. S. (2006b). Hedgehog Signaling and Bmi-1 Regulate Self-renewal of Normal and Malignant Human Mammary Stem Cells. *Cancer Research*, *66*(12), 6063–6071. <https://doi.org/10.1158/0008-5472.can-06-0054>
- Liu, X., Liu, X., Shi, Q., Fan, X., & Qi, K. (2021b). Association of telomere length and telomerase methylation with n-3 fatty acids in preschool children with obesity. *BMC Pediatrics*, *21*(1), 24. <https://doi.org/10.1186/s12887-020-02487-x>
- Liu, X., Shi, Q., Fan, X., Chen, H., Chen, N., Zhao, Y., & Qi, K. (2022). Associations of Maternal Polyunsaturated Fatty Acids With Telomere Length in the Cord Blood and Placenta in Chinese Population. *Frontiers in Nutrition*, *8*, 779306. <https://doi.org/10.3389/fnut.2021.779306>
- Liu, Y., Han, Y., Zhou, L., Pan, X., Sun, X., Liu, Y., Liang, M., Qin, J., Lu, Y., & Liu, P. (2020d). A comprehensive evaluation of computational tools to identify differential methylation regions using RRBS data. *Genomics*, *112*(6), 4567–4576. <https://doi.org/10.1016/j.ygeno.2020.07.032>
- Liu, Y., Wang, G., Zhang, J., Chen, X., Xu, H., Heng, G., Chen, J., Zhao, Y., Li, J., Ni, Y., Zhang, Y., Shan, J., & Qian, C. (2021c). CD9, a potential leukemia stem cell marker, regulates drug resistance and leukemia development in acute myeloid leukemia. *Stem Cell Research & Therapy*, *12*(1), 86. <https://doi.org/10.1186/s13287-021-02155-6>
- Lizardi, P. M., Yan, Q., & Wajapeyee, N. (2016). DNA Bisulfite Sequencing for Single-Nucleotide-Resolution DNA Methylation Detection. *Cold Spring Harbor Protocols*, *2017*(11), pdb.prot094839. <https://doi.org/10.1101/pdb.prot094839>
- Locatelli, M. A., Aftimos, P., Dees, E. C., LoRusso, P. M., Pegram, M. D., Awada, A., Huang, B., Cesari, R., Jiang, Y., Shaik, M. N., Kern, K. A., & Curigliano, G. (2016). Phase I study of the gamma secretase inhibitor PF-03084014 in combination with docetaxel in patients with advanced triple-negative breast cancer. *Oncotarget*, *8*(2), 2320–2328. <https://doi.org/10.18632/oncotarget.13727>
- Loewer, S., Cabili, M. N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I. H., Garber, M., Curran, M., Onder, T., Agarwal, S., Manos, P. D., Datta, S., Lander, E. S., Schlaeger, T. M., Daley, G. Q., & Rinn, J. L. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics*, *42*(12), 1113–1117. <https://doi.org/10.1038/ng.710>
- Lombardo, Y., Giorgio, A. d., Coombes, C. R., Stebbing, J., & Castellano, L. (2015). Mammosphere Formation Assay from Human Breast Cancer Tissues and Cell Lines. *Journal of Visualized Experiments*, (97). <https://doi.org/10.3791/52671>
- Longley, D., & Johnston, P. (2005). Molecular mechanisms of drug resistance. *The Journal of Pathology*, *205*(2), 275–292. <https://doi.org/10.1002/path.1706>
- Louie, E., Nik, S., Chen, J.-s., Schmidt, M., Song, B., Pacson, C., Chen, X. F., Park, S., Ju, J., & Chen, E. I. (2010). Identification of a stem-like cell population by exposing metastatic breast cancer cell lines to repetitive cycles of hypoxia and reoxygenation. *Breast Cancer Research : BCR*, *12*(6), R94–R94. <https://doi.org/10.1186/bcr2773>
- Lu, D.-g., Ma, Y.-m., Zhu, A.-j., & Han, Y.-w. (2016). An early biomarker and potential therapeutic target of RUNX 3 hypermethylation in breast cancer, a system review and meta-analysis. *Oncotarget*, *8*(13), 22166–22174. <https://doi.org/10.18632/oncotarget.13125>
- Lu, H., Clauser, K. R., Tam, W. L., Fröse, J., Ye, X., Eaton, E. N., Reinhardt, F., Donnem, V. S., Bhargava, R., Carr, S. A., & Weinberg, R. A. (2014a). A breast cancer stem cell niche supported by juxtacrine signalling from monocytes and macrophages. *Nature Cell Biology*, *16*(11), 1105–1117. <https://doi.org/10.1038/ncb3041>
- Lu, H., Samanta, D., Xiang, L., Zhang, H., Hu, H., Chen, I., Bullen, J. W., & Semenza, G. L. (2015). Chemotherapy triggers HIF-1-dependent glutathione synthesis and copper chelation that induces the breast cancer stem cell phenotype. *Proceedings of the National Academy of Sciences*, *112*(33), E4600–E4609. <https://doi.org/10.1073/pnas.1513433112>
- Lu, P., Weaver, V. M., & Werb, Z. (2012). The extracellular matrix: A dynamic niche in cancer progression. *The Journal of Cell Biology*, *196*(4), 395–406. <https://doi.org/10.1083/jcb.201102147>
- Lu, Y., Lu, J., Li, X., Zhu, H., Fan, X., Zhu, S., Wang, Y., Guo, Q., Wang, L., Huang, Y., Zhu, M., & Wang, Z. (2014b). MiR-200a inhibits epithelial-mesenchymal transition of pancreatic cancer stem cell. *BMC Cancer*, *14*(1), 85–85. <https://doi.org/10.1186/1471-2407-14-85>
- Lucero, M., Thind, J., Sandoval, J., Senaati, S., Jimenez, B., & Kandpal, R. P. (2020). Stem-like Cells from Invasive Breast Carcinoma Cell Line MDA-MB-231 Express a Distinct Set of Eph Receptors and Ephrin Ligands. *Cancer Genomics - Proteomics*, *17*(6), 729–738. <https://doi.org/10.21873/cgp.20227>
- Lytle, N. K., Barber, A. G., & Reya, T. (2018). Stem cell fate in cancer growth, progression and therapy resistance. *Nature Reviews Cancer*, *18*(11), 669–680. <https://doi.org/10.1038/s41568-018-0056-x>

M

- Ma, Y., Shen, N., Wicha, M. S., & Luo, M. (2021). The Roles of the Let-7 Family of MicroRNAs in the Regulation of Cancer Stemness. *Cells*, *10*(9), 2415. <https://doi.org/10.3390/cells10092415>

- Maekawa, M., Sugano, K., Kashiwabara, H., Ushiyama, M., Fujita, S., Yoshimori, M., & Kakizoe, T. (1999). DNA Methylation Analysis Using Bisulfite Treatment and PCR–Single-Strand Conformation Polymorphism in Colorectal Cancer Showing Microsatellite Instability. *Biochemical and Biophysical Research Communications*, 262(3), 671–676. <https://doi.org/10.1006/bbrc.1999.1230>
- Makki, J. (2015). Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clinical Medicine Insights. Pathology*, 8, 23–31. <https://doi.org/10.4137/cpath.s31563>
- Makki, J., Myint, O., Wynn, A. A., Samsudin, A. T., & John, D. V. (2015). Expression Distribution of Cancer Stem Cells, Epithelial to Mesenchymal Transition, and Telomerase Activity in Breast Cancer and Their Association with Clinicopathologic Characteristics. *Clinical Medicine Insights. Pathology*, 8, 1–16. <https://doi.org/10.4137/cpath.s19615>
- Malhotra, G. K., Zhao, X., Band, H., & Band, V. (2010). Histological, molecular and functional subtypes of breast cancers. *Cancer Biology & Therapy*, 10(10), 955–960. <https://doi.org/10.4161/cbt.10.10.13879>
- Mallona, I., Díez-Villanueva, A., & Peinado, M. A. (2014). Methylation plotter: a web tool for dynamic visualization of DNA methylation data. *Source Code for Biology and Medicine*, 9(1), 11–11. <https://doi.org/10.1186/1751-0473-9-11>
- Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell*, 133(4), 704–715. <https://doi.org/10.1016/j.cell.2008.03.027>
- Mansego, M. L., Milagro, F. I., Campión, J., & Martínez, J. A. (2013). Techniques of DNA Methylation Analysis with Nutritional Applications. *Lifestyle Genomics*, 6(2), 83–96. <https://doi.org/10.1159/000350749>
- Mansour, S. F., & Atwa, M. M. (2015). Clinicopathological Significance of CD133 and ALDH1 Cancer Stem Cell Marker Expression in Invasive Ductal Breast Carcinoma. *Asian Pacific Journal of Cancer Prevention*, 16(17), 7491–7496. <https://doi.org/10.7314/apjcp.2015.16.17.7491>
- Marcato, P., Dean, C. A., Giacomantonio, C. A., & Lee, P. W. K. (2011a). Aldehyde dehydrogenase: Its role as a cancer stem cell marker comes down to the specific isoform. *Cell Cycle*, 10(9), 1378–1384. <https://doi.org/10.4161/cc.10.9.15486>
- Marcato, P., Dean, C. A., Pan, D., Araslanova, R., Gillis, M., Joshi, M., Helyer, L., Pan, L., Leidal, A., Gujar, S., Giacomantonio, C. A., & Lee, P. W. K. (2011b). Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells Is Primarily Due To Isoform ALDH1A3 and Its Expression Is Predictive of Metastasis. *STEM CELLS*, 29(1), 32–45. <https://doi.org/10.1002/stem.563>
- Marcucci, F., Caserta, C. A., Romeo, E., & Rumio, C. (2019). Antibody-Drug Conjugates (ADC) Against Cancer Stem-Like Cells (CSC)—Is There Still Room for Optimism? *Frontiers in Oncology*, 9, 167. <https://doi.org/10.3389/fonc.2019.00167>
- Marcucci, F., & Corti, A. (2012). Improving drug penetration to curb tumor drug resistance. *Drug Discovery Today*, 17(19–20), 1139–1146. <https://doi.org/10.1016/j.drudis.2012.06.004>
- Marjanovic, N. D., Weinberg, R. A., & Chaffer, C. L. (2013). Cell Plasticity and Heterogeneity in Cancer. *Clinical Chemistry*, 59(1), 168–179. <https://doi.org/10.1373/clinchem.2012.184655>
- Markopoulos, G. S., Roupakia, E., Marcu, K. B., & Kolettas, E. (2019). Epigenetic Regulation of Inflammatory Cytokine-Induced Epithelial-To-Mesenchymal Cell Transition and Cancer Stem Cell Generation. *Cells*, 8(10), 1143. <https://doi.org/10.3390/cells8101143>
- Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K., Miller, R. A., Digles, D., Lopes, E. N., Ehrhart, F., Dupuis, L. J., Winkers, L. A., Coort, S. L., Willighagen, E. L., Evelo, C. T., Pico, A. R., & Kutmon, M. (2020). WikiPathways: connecting communities. *Nucleic Acids Research*, 49(D1), D613–D621. <https://doi.org/10.1093/nar/gkaa1024>
- Martín, B., Pappa, S., Díez-Villanueva, A., Mallona, I., Custodio, J., Barrero, M. J., Peinado, M. A., & Jordà, M. (2020). Tissue and cancer-specific expression of DIXF is epigenetically mediated by an Alu repeat. *Epigenetics*, 15(6–7), 1–15. <https://doi.org/10.1080/15592294.2020.1722398>
- Martisova, A., Holcakova, J., Izadi, N., Sebuyoya, R., Hrstka, R., & Bartosik, M. (2021). DNA Methylation in Solid Tumors: Functions and Methods of Detection. *International Journal of Molecular Sciences*, 22(8), 4247. <https://doi.org/10.3390/ijms22084247>
- Masoumi, J., Jafarzadeh, A., Abdolalizadeh, J., Khan, H., Philippe, J., Mirzaei, H., & Mirzaei, H. R. (2021). Cancer stem cell-targeted chimeric antigen receptor (CAR)-T cell therapy: Challenges and prospects. *Acta Pharmaceutica Sinica B*, 11(7), 1721–1739. <https://doi.org/10.1016/j.apsb.2020.12.015>
- Masui, T., Ota, I., Yook, J.-I., Mikani, S., Yane, K., Yamanaka, T., & Hosoi, H. (2014). Snail-induced epithelial-mesenchymal transition promotes cancer stem cell-like phenotype in head and neck cancer cells. *International Journal of Oncology*, 44(3), 693–699. <https://doi.org/10.3892/ijo.2013.2225>
- Matsui, W. H. (2016). Cancer stem cell signaling pathways. *Medicine*, 95(Suppl 1), S8–S19. <https://doi.org/10.1097/md.0000000000004765>

- Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., ... Costello, J. F. (2010). Conserved Role of Intragenic DNA Methylation in Regulating Alternative Promoters. *Nature*, *466*(7303), 253–257. <https://doi.org/10.1038/nature09165>
- McDonald, L. E., & Kay, G. F. (1997). Methylation Analysis Using Bisulfite Genomic Sequencing: Application to Small Numbers of Intact Cells. *BioTechniques*, *22*(2), 272–274. <https://doi.org/10.2144/97222bm16>
- McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, *168*(4), 613–628. <https://doi.org/10.1016/j.cell.2017.01.018>
- Mehdipour, P., Murphy, T., & Carvalho, D. D. D. (2020). The role of DNA-demethylating agents in cancer therapy. *Pharmacology & Therapeutics*, *205*, 107416. <https://doi.org/10.1016/j.pharmthera.2019.107416>
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, *33*(18), 5868–5877. <https://doi.org/10.1093/nar/gki901>
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. <https://doi.org/10.1038/nature07107>
- Meldolesi, J. (2022). Cancer Stem Cells and Their Vesicles, Together with Other Stem and Non-Stem Cells, Govern Critical Cancer Processes: Perspectives for Medical Development. *International Journal of Molecular Sciences*, *23*(2), 625. <https://doi.org/10.3390/ijms23020625>
- Melnikov, A. A., Gartenhaus, R. B., Levenson, A. S., Motchoulskaia, N. A., & (Chernokhvostov), V. V. L. (2005). MSRE-PCR for analysis of gene-specific DNA methylation. *Nucleic Acids Research*, *33*(10), e93–e93. <https://doi.org/10.1093/nar/gni092>
- Melzer, C., Ohe, J. v. d., Lehnert, H., Ungefroren, H., & Hass, R. (2017). Cancer stem cell niche models and contribution by mesenchymal stroma/stem cells. *Molecular Cancer*, *16*(1), 28. <https://doi.org/10.1186/s12943-017-0595-x>
- Meng, H., Cao, Y., Qin, J., Song, X., Zhang, Q., Shi, Y., & Cao, L. (2015). DNA Methylation, Its Mediators and Genome Integrity. *International Journal of Biological Sciences*, *11*(5), 604–617. <https://doi.org/10.7150/ijbs.11218>
- Mertins, S. D. (2014). Cancer stem cells. *Anti-Cancer Drugs*, *25*(4), 353–367. <https://doi.org/10.1097/cad.0000000000000075>
- Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems. (2007). <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.616.1849>
- Michealraj, K. A., Kumar, S. A., Kim, L. J., Cavalli, F. M., Przelicki, D., Wojcik, J. B., Delaidelli, A., Bajic, A., Saulnier, O., MacLeod, G., Vellanki, R. N., Vladioiu, M. C., Guilhamon, P., Ong, W., Lee, J. J., Jiang, Y., Holgado, B. L., Rasnitsyn, A., Malik, A. A., ... Taylor, M. D. (2020). Metabolic Regulation of the Epigenome Drives Lethal Infantile Ependymoma. *Cell*, *181*(6), 1329–1345.e24. <https://doi.org/10.1016/j.cell.2020.04.047>
- Michor, F., & Polyak, K. (2010). The Origins and Implications of Intratumor Heterogeneity. *Cancer Prevention Research*, *3*(11), 1361–1364. <https://doi.org/10.1158/1940-6207.capr-10-0234>
- Mikeska, T., Candiloro, I. L., & Dobrovic, A. (2010). The implications of heterogeneous DNA methylation for the accurate quantification of methylation. *Epigenomics*, *2*(4), 561–573. <https://doi.org/10.2217/epi.10.32>
- Milosevic, V., Kopecka, J., Salaroglio, I. C., Libener, R., Napoli, F., Izzo, S., Orecchia, S., Ananthanarayanan, P., Bironzo, P., Grosso, F., Tabbò, F., Comunanza, V., Alexa-Stratulat, T., Bussolino, F., Righi, L., Novello, S., Scagliotti, G. V., & Riganti, C. (2020). Wnt/IL-1 β /IL-8 autocrine circuitries control chemoresistance in mesothelioma initiating cells by inducing ABCB5. *International Journal of Cancer*, *146*(1), 192–207. <https://doi.org/10.1002/ijc.32419>
- Minchinton, A. I., & Tannock, I. F. (2006). Drug penetration in solid tumours. *Nature Reviews Cancer*, *6*(8), 583–592. <https://doi.org/10.1038/nrc1893>
- Miousse, I. R., Kutanzi, K. R., & Koturbash, I. (2017). Effects of ionizing radiation on DNA methylation: from experimental biology to clinical applications. *International Journal of Radiation Biology*, *93*(5), 1–39. <https://doi.org/10.1080/09553002.2017.1287454>
- Miousse, I. R., Shao, L., Chang, J., Feng, W., Wang, Y., Allen, A. R., Turner, J., Stewart, B., Raber, J., Zhou, D., & Koturbash, I. (2014). Exposure to Low-Dose 56Fe-Ion Radiation Induces Long-Term Epigenetic Alterations in Mouse Bone Marrow Hematopoietic Progenitor and Stem Cells. *Radiation Research*, *182*(1), 92–101. <https://doi.org/10.1667/rr13580.1>
- Mitchell, C., Schneper, L. M., & Notterman, D. A. (2016). DNA methylation, early life environment, and health outcomes. *Pediatric Research*, *79*(1-2), 212–219. <https://doi.org/10.1038/pr.2015.193>
- Mohammed, M. K., Shao, C., Wang, J., Wei, Q., Wang, X., Collier, Z., Tang, S., Liu, H., Zhang, F., Huang, J., Guo, D., Lu, M., Liu, F., Liu, J., Ma, C., Shi, L. L., Athiviraham, A., He, T.-C., & Lee, M. J. (2016). Wnt/ β -catenin signaling plays an

- ever-expanding role in stem cell self-renewal, tumorigenesis and cancer chemoresistance. *Genes & Diseases*, 3(1), 11–40. <https://doi.org/10.1016/j.gendis.2015.12.004>
- Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., & Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Molecular Cell*, 30(6), 755–766. <https://doi.org/10.1016/j.molcel.2008.05.007>
- Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1), 23–38. <https://doi.org/10.1038/npp.2012.112>
- Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., & Puisieux, A. (2008). Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. *PLoS ONE*, 3(8), e2888. <https://doi.org/10.1371/journal.pone.0002888>
- Morgan, A., LeGresley, S., & Fischer, C. (2020). Remodeler Catalyzed Nucleosome Repositioning: Influence of Structure and Stability. *International Journal of Molecular Sciences*, 22(1), 76. <https://doi.org/10.3390/ijms22010076>
- Moschny, N., Jahn, K., Bajbouj, M., Maier, H. B., Ballmaier, M., Khan, A. Q., Pollak, C., Bleich, S., Frieling, H., & Neyazi, A. (2020). DNA Methylation of the t-PA Gene Differs Between Various Immune Cell Subtypes Isolated From Depressed Patients Receiving Electroconvulsive Therapy. *Frontiers in Psychiatry*, 11, 571. <https://doi.org/10.3389/fpsy.2020.00571>
- Moskalev, E. A., Zavgorodnij, M. G., Majorova, S. P., Vorobjev, I. A., Jandaghi, P., Bure, I. V., & Hoheisel, J. D. (2011). Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Research*, 39(11), e77–e77. <https://doi.org/10.1093/nar/gkr213>
- Movahed, Z. G., Rastegari-Pouyani, M., Mohammadi, M. h., & Mansouri, K. (2019). Cancer cells change their glucose metabolism to overcome increased ROS: One step from cancer cell to cancer stem cell? *Biomedicine & Pharmacotherapy*, 112, 108690. <https://doi.org/10.1016/j.biopha.2019.108690>
- Mu, R., Zou, Y.-K., Tu, K., Wang, D.-B., Tang, D., Yu, Z., & Zhao, L. (2021). Hypoxia Promotes Pancreatic Cancer Cell Dedifferentiation to Stem-Like Cell Phenotypes With High Tumorigenic Potential by the HIF-1 α /Notch Signaling Pathway. *Pancreas*, 50(5), 756–765. <https://doi.org/10.1097/mpa.0000000000001828>
- Mu, X., Español-Suñer, R., Mederacke, I., Affò, S., Manco, R., Sempoux, C., Lemaigre, F. P., Adili, A., Yuan, D., Weber, A., Unger, K., Heikenwälder, M., Leclercq, I. A., & Schwabe, R. F. (2015). Hepatocellular carcinoma originates from hepatocytes and not from the progenitor/biliary compartment. *Journal of Clinical Investigation*, 125(10), 3891–3903. <https://doi.org/10.1172/jci77995>
- Mühlich, J., Bajanowski, T., Rickert, C. H., Roggendorf, W., Würthwein, G., Jürgens, H., & Frühwald, M. C. (2007). Frequent but borderline methylation of p16INK4a and TIMP3 in medulloblastoma and sPNET revealed by quantitative analyses. *Journal of Neuro-Oncology*, 83(1), 17–29. <https://doi.org/10.1007/s11060-006-9309-8>
- Muñoz, P., Iliou, M. S., & Esteller, M. (2012). Epigenetic alterations involved in cancer stem cell reprogramming. *Molecular Oncology*, 6(6), 620–636. <https://doi.org/10.1016/j.molonc.2012.10.006>
- Murata, K., Saga, R., Monzen, S., Tsuruga, E., Hasegawa, K., & Hosokawa, Y. (2019). Understanding the mechanism underlying the acquisition of radioresistance in human prostate cancer cells. *Oncology Letters*, 17(6), 5830–5838. <https://doi.org/10.3892/ol.2019.10219>
- Mutschelknaus, L., Peters, C., Winkler, K., Yentrapalli, R., Heider, T., Atkinson, M. J., & Moertl, S. (2016). Exosomes Derived from Squamous Head and Neck Cancer Promote Cell Survival after Ionizing Radiation. *PLoS ONE*, 11(3), e0152213. <https://doi.org/10.1371/journal.pone.0152213>
- Myöhänen, S., Wahlfors, J., & Jänne, J. (1994). Automated fluorescent genomic sequencing as applied to the methylation analysis of the human ornithine decarboxylase gene. *DNA Sequence*, 5(1), 1–8. <https://doi.org/10.3109/10425179409039698>

N

- Nabholtz, J.-M., & Gligorov, J. (2005). The role of taxanes in the treatment of breast cancer. *Expert Opinion on Pharmacotherapy*, 6(7), 1073–1094. <https://doi.org/10.1517/14656566.6.7.1073>
- Nagane, Y., Utsugisawa, K., & Tohgi, H. (2000). PCR amplification in bisulfite methylcytosine mapping in the GC-rich promoter region of amyloid precursor protein gene in autopsy human brain. *Brain Research Protocols*, 5(2), 167–171. [https://doi.org/10.1016/s1385-299x\(00\)00008-8](https://doi.org/10.1016/s1385-299x(00)00008-8)
- Nagare, R. P., Sneha, S., Sidhanth, C., Roopa, S., Murhekar, K., Shirley, S., Swaminathan, R., Sridevi, V., & Ganesan, T. S. (2020). Expression of cancer stem cell markers CD24, EPHA1 and CD9 and their correlation with clinical outcome in epithelial ovarian tumours. *Cancer Biomarkers*, 28(3), 397–408. <https://doi.org/10.3233/cbm-201463>
- Nair, N., Calle, A. S., Zahra, M. H., Prieto-Vila, M., Oo, A. K. K., Hurley, L., Vaidyanath, A., Seno, A., Masuda, J., Iwasaki, Y., Tanaka, H., Kasai, T., & Seno, M. (2017). A cancer stem cell model as the point of origin of cancer-associated fibroblasts in tumor microenvironment. *Scientific Reports*, 7(1), 6838. <https://doi.org/10.1038/s41598-017-07144-5>

- Najafi, M., Farhood, B., & Mortezaee, K. (2018). Cancer stem cells (CSCs) in cancer progression and therapy. *Journal of Cellular Physiology*, 234(6), 8381–8395. <https://doi.org/10.1002/jcp.27740>
- Najafi, M., Mortezaee, K., & Majidpoor, J. (2019). Cancer stem cell (CSC) resistance drivers. *Life Sciences*, 234, 116781. <https://doi.org/10.1016/j.lfs.2019.116781>
- Nakano, M., Kikushige, Y., Miyawaki, K., Kunisaki, Y., Mizuno, S., Takenaka, K., Tamura, S., Okumura, Y., Ito, M., Ariyama, H., Kusaba, H., Nakamura, M., Maeda, T., Baba, E., & Akashi, K. (2018). Dedifferentiation process driven by TGF-beta signaling enhances stem cell properties in human colorectal cancer. *Oncogene*, 38(6), 780–793. <https://doi.org/10.1038/s41388-018-0480-0>
- Nam, A. S., Chaligne, R., & Landau, D. A. (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Reviews Genetics*, 22(1), 3–18. <https://doi.org/10.1038/s41576-020-0265-5>
- Nan, X., Ng, H.-H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), 386–389. <https://doi.org/10.1038/30764>
- Nassar, D., & Blanpain, C. (2016). Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annual Review of Pathology: Mechanisms of Disease*, 11(1), 47–76. <https://doi.org/10.1146/annurev-pathol-012615-044438>
- National Cancer Institute (US). (n.d.). <https://www.cancer.gov/>
- Natsume, A., Ito, M., Katsushima, K., Ohka, F., Hatanaka, A., Shinjo, K., Sato, S., Takahashi, S., Ishikawa, Y., Takeuchi, I., Shimogawa, H., Uesugi, M., Okano, H., Kim, S. U., Wakabayashi, T., Issa, J.-P. J., Sekido, Y., & Kondo, Y. (2013). Chromatin Regulator PRC2 Is a Key Regulator of Epigenetic Plasticity in Glioblastoma. *Cancer Research*, 73(14), 4559–4570. <https://doi.org/10.1158/0008-5472.can-13-0109>
- Nazio, F., Bordi, M., Cianfanelli, V., Locatelli, F., & Cecconi, F. (2019). Autophagy and cancer stem cells: molecular mechanisms and therapeutic applications. *Cell Death and Differentiation*, 26(4), 690–702. <https://doi.org/10.1038/s41418-019-0292-y>
- Ni, S.-J., Zhao, L.-Q., Wang, X.-F., Wu, Z.-H., Hua, R.-X., Wan, C.-H., Zhang, J.-Y., Zhang, X.-W., Huang, M.-Z., Gan, L., Sun, H.-L., Dimri, G. P., & Guo, W.-J. (2018). CBX7 regulates stem cell-like properties of gastric cancer cells via p16 and AKT-NF- κ B-miR-21 pathways. *Journal of Hematology & Oncology*, 11(1), 17. <https://doi.org/10.1186/s13045-018-0562-z>
- Niklaus, N. J., Tokarchuk, I., Zbinden, M., Schläfli, A. M., Maycotte, P., & Tschan, M. P. (2021). The Multifaceted Functions of Autophagy in Breast Cancer Development and Treatment. *Cells*, 10(6), 1447. <https://doi.org/10.3390/cells10061447>
- Nishida, H., Yamazaki, H., Yamada, T., Iwata, S., Dang, N. H., Inukai, T., Sugita, K., Ikeda, Y., & Morimoto, C. (2009). CD9 correlates with cancer stem cell potentials in human B-acute lymphoblastic leukemia cells. *Biochemical and Biophysical Research Communications*, 382(1), 57–62. <https://doi.org/10.1016/j.bbrc.2009.02.123>
- Nishina, S.-I., Shiraha, H., Nakanishi, Y., Tanaka, S., Matsubara, M., Takaoka, N., Uemura, M., Horiguchi, S., Kataoka, J., Iwamuro, M., Yagi, T., & Yamamoto, K. (2011). Restored expression of the tumor suppressor gene RUNX3 reduces cancer stem cells in hepatocellular carcinoma by suppressing Jagged1-Notch signaling. *Oncology Reports*, 26(3), 523–31. <https://doi.org/10.3892/or.2011.1336>
- Nishino, K., & Umezawa, A. (2016). DNA methylation dynamics in human induced pluripotent stem cells. *Human Cell*, 29(3), 97–100. <https://doi.org/10.1007/s13577-016-0139-5>
- Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations. *Science*, 194(4260), 23–28. <https://doi.org/10.1126/science.959840>
- Núñez, C., Capelo, J. L., Igrejas, G., Alfonso, A., Botana, L. M., & Lodeiro, C. (2016). An overview of the effective combination therapies for the treatment of breast cancer. *Biomaterials*, 97, 34–50. <https://doi.org/10.1016/j.biomaterials.2016.04.027>
- Nussbaumer, S., Bonnabry, P., Veuthey, J.-L., & Fleury-Souverain, S. (2011). Analysis of anticancer drugs: A review. *Talanta*, 85(5), 2265–2289. <https://doi.org/10.1016/j.talanta.2011.08.034>
- Nygren, A. O. H., Ameziane, N., Duarte, H. M. B., Vijzelaar, R. N. C. P., Waisfisz, Q., Hess, C. J., Schouten, J. P., & Errami, A. (2005). Methylation-Specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Research*, 33(14), e128–e128. <https://doi.org/10.1093/nar/gni127>
- O**
- Oakeley, E. J. (1999). DNA methylation analysis a review of current methodologies. *Pharmacology & Therapeutics*, 84(3), 389–400. [https://doi.org/10.1016/s0163-7258\(99\)00043-1](https://doi.org/10.1016/s0163-7258(99)00043-1)
- O'Brien, C. A., Pollett, A., Gallinger, S., & Dick, J. E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*, 445(7123), 106–110. <https://doi.org/10.1038/nature05372>

- O'Brien, C. A., Kreso, A., & Jamieson, C. H. (2010). Cancer Stem Cells and Self-renewal. *Clinical Cancer Research*, 16(12), 3113–3120. <https://doi.org/10.1158/1078-0432.ccr-09-2824>
- O'Connor, R. (2007). The pharmacology of cancer resistance. *Anticancer research*, 27(3A), 1267–72.
- O'Connor, T. P., & Crystal, R. G. (2006). Genetic medicines: treatment strategies for hereditary disorders. *Nature Reviews Genetics*, 7(4), 261–276. <https://doi.org/10.1038/nrg1829>
- Oda, M., Glass, J. L., Thompson, R. F., Mo, Y., Olivier, E. N., Figueroa, M. E., Selzer, R. R., Richmond, T. A., Zhang, X., Dannenberg, L., Green, R. D., Melnick, A., Hatchwell, E., Bouhassira, E. E., Verma, A., Suzuki, M., & Grealley, J. M. (2009). High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Research*, 37(12), 3829–3839. <https://doi.org/10.1093/nar/gkp260>
- Oikawa, T. (2016). Cancer Stem cells and their cellular origins in primary liver and biliary tract cancers. *Hepatology*, 64(2), 645–651. <https://doi.org/10.1002/hep.28485>
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, 99(3), 247–257. [https://doi.org/10.1016/s0092-8674\(00\)81656-6](https://doi.org/10.1016/s0092-8674(00)81656-6)
- Olariu, V., Lökvist, C., & Sneppen, K. (2016). Nanog, Oct4 and Tet1 interplay in establishing pluripotency. *Scientific Reports*, 6(1), 25438. <https://doi.org/10.1038/srep25438>
- Olek, A., Oswald, J., & Walter, J. (1996). A Modified and Improved Method for Bisulphite Based Cytosine Methylation Analysis. *Nucleic Acids Research*, 24(24), 5064–5066. <https://doi.org/10.1093/nar/24.24.5064>
- Olmez, I., Shen, W., McDonald, H., & Ozpolat, B. (2015). Dedifferentiation of patient-derived glioblastoma multiforme cell lines results in a cancer stem cell-like state with mitogen-independent growth. *Journal of Cellular and Molecular Medicine*, 19(6), 1262–1272. <https://doi.org/10.1111/jcmm.12479>
- Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrrens, R. V., Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, 19(1), 33. <https://doi.org/10.1186/s13059-018-1408-2>
- Ordway, J., Bedell, J., Citek, R., Nunberg, A., Garrido, A., Kendall, R., Stevens, J., Cao, D., Doerge, R., Korshunova, Y., Holemon, H., McPherson, J., Lakey, N., Leon, J., Martienssen, R., & Jeddloh, J. (2006). Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis*, 27(12), 2409–2423. <https://doi.org/10.1093/carcin/bgl161>
- Orian-Rousseau, V. (2015). CD44 Acts as a Signaling Platform Controlling Tumor Progression and Metastasis. *Frontiers in Immunology*, 6, 154. <https://doi.org/10.3389/fimmu.2015.00154>
- Orzan, F., Pellegatta, S., Poliani, P. L., Pisati, F., Caldera, V., Menghi, F., Kapetis, D., Marras, C., Schiffer, D., & Finocchiaro, G. (2011). Enhancer of Zeste 2 (EZH2) is up-regulated in malignant gliomas and in glioma stem-like cells. *Neuropathology and Applied Neurobiology*, 37(4), 381–394. <https://doi.org/10.1111/j.1365-2990.2010.01132.x>
- Osborne, C., Wilson, P., & Tripathy, D. (2004). Oncogenes and Tumor Suppressor Genes in Breast Cancer: Potential Diagnostic and Therapeutic Applications. *The Oncologist*, 9(4), 361–377. <https://doi.org/10.1634/theoncologist.9-4-361>
- Oshima, N., Yamada, Y., Nagayama, S., Kawada, K., Hasegawa, S., Okabe, H., Sakai, Y., & Aoi, T. (2014). Induction of Cancer Stem Cell Properties in Colon Cancer Cells by Defined Factors. *PLoS ONE*, 9(7), e101735. <https://doi.org/10.1371/journal.pone.0101735>
- Osswald, M., Blaes, J., Liao, Y., Solecki, G., Gömmel, M., Berghoff, A. S., Salphati, L., Wallin, J. J., Phillips, H. S., Wick, W., & Winkler, F. (2016). Impact of Blood–Brain Barrier Integrity on Tumor Growth and Therapy Response in Brain Metastases. *Clinical Cancer Research*, 22(24), 6078–6087. <https://doi.org/10.1158/1078-0432.ccr-16-1327>

P

- Page, D. L., & Rogers, L. W. (1992). Combined histologic and cytologic criteria for the diagnosis of mammary atypical ductal hyperplasia. *Human Pathology*, 23(10), 1095–1097. [https://doi.org/10.1016/0046-8177\(92\)90026-y](https://doi.org/10.1016/0046-8177(92)90026-y)
- Pajares, M. J., Palanca-Ballester, C., Urtasun, R., Alemany-Cosme, E., Lahoz, A., & Sandoval, J. (2021). Methods for analysis of specific DNA methylation status. *Methods*, 187, 3–12. <https://doi.org/10.1016/j.ymeth.2020.06.021>
- Pajonk, F., Vlashi, E., & McBride, W. H. (2010). Radiation Resistance of Cancer Stem Cells: The 4 R's of Radiobiology Revisited. *STEM CELLS*, 28(4), 639–648. <https://doi.org/10.1002/stem.318>
- Pan, Y., Chen, J., Tao, L., Zhang, K., Wang, R., Chu, X., & Chen, L. (2017). Long noncoding RNA ROR regulates chemoresistance in docetaxel-resistant lung adenocarcinoma cells via epithelial mesenchymal transition pathway. *Oncotarget*, 8(20), 33144–33158. <https://doi.org/10.18632/oncotarget.16562>
- Pang, D., Zhao, Y., Xue, W., Shan, M., Chen, Y., Zhang, Y., Zhang, G., Liu, F., Li, D., & Yang, Y. (2012). Methylation profiles of the BRCA1 promoter in hereditary and sporadic breast cancer among Han Chinese. *Medical Oncology*, 29(3), 1561–1568. <https://doi.org/10.1007/s12032-011-0100-0>

- Parise, C. A., & Caggiano, V. (2014). Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *Journal of Cancer Epidemiology*, 2014, 469251. <https://doi.org/10.1155/2014/469251>
- Park, H., Cho, B., & Kim, J. (2020). Rad50 mediates DNA demethylation to establish pluripotent reprogramming. *Experimental & Molecular Medicine*, 52(7), 1116–1127. <https://doi.org/10.1038/s12276-020-0467-0>
- Park, S.-J., Shim, J. W., Park, H. S., Eum, D.-Y., Park, M.-T., Yi, J. M., Choi, S. H., Kim, S. D., Son, T. G., Lu, W., Kim, N. D., Yang, K., & Heo, K. (2016). MacroH2A1 downregulation enhances the stem-like properties of bladder cancer cells by transactivation of Lin28B. *Oncogene*, 35(10), 1292–1301. <https://doi.org/10.1038/onc.2015.187>
- Park, Y., Figueroa, M. E., Rozek, L. S., & Sartor, M. A. (2014). MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics*, 30(17), 2414–2422. <https://doi.org/10.1093/bioinformatics/btu339>
- Parker, L. T., Deng, Q., Zakeri, H., Carlson, C., Nickerson, D. A., & Kwok, P. Y. (1995). Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry. *BioTechniques*, 19(1), 116–21.
- Parnell, C., & Woll, P. J. (2005). Principles of cancer chemotherapy. *The Foundation Years*, 1(2), 18–22. [https://doi.org/10.1053/s1744-1889\(06\)70015-9](https://doi.org/10.1053/s1744-1889(06)70015-9)
- Parrish, R. R., Day, J. J., & Lubin, F. D. (2012). Direct Bisulfite Sequencing for Examination of DNA Methylation with Gene and Nucleotide Resolution from Brain Tissues. *Current Protocols in Neuroscience*, 60(1), 7.24.1–7.24.12. <https://doi.org/10.1002/0471142301.ns0724s60>
- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Keymeulen, A. V., Brown, D., Moers, V., Lemaire, S., Clercq, S. D., Minguijón, E., Balsat, C., Sokolow, Y., Dubois, C., Cock, F. D., Scozzaro, S., Sopena, F., Lanas, A., ... Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nature*, 556(7702), 463–468. <https://doi.org/10.1038/s41586-018-0040-3>
- Pathania, R., Ramachandran, S., Elangovan, S., Padia, R., Yang, P., Cinghu, S., Veeranan-Karmegam, R., Arjunan, P., Gnana-Prakasam, J. P., Fulzele, S., Pei, L., Chang, C.-S., Choi, H., Shi, H., Manicassamy, S., Prasad, P. D., Sharma, S., Ganapathy, V., Jothi, R., & Thangaraju, M. (2015). DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. *Nature communications*, 6(1), 6910–6910. <https://doi.org/10.1038/ncomms7910>
- Patil, S., Steuber, B., Kopp, W., Kari, V., Urbach, L., Wang, X., Küffer, S., Bohnenberger, H., Spyropoulou, D., Zhang, Z., Versemann, L., Bösherz, M. S., Brunner, M., Gaedcke, J., Ströbel, P., Zhang, J.-S., Neesse, A., Ellenrieder, V., Singh, S. K., ... Hessmann, E. (2020). EZH2 Regulates Pancreatic Cancer Subtype Identity and Tumor Progression via Transcriptional Repression of GATA6. *Cancer Research*, 80(21), 4620–4632. <https://doi.org/10.1158/0008-5472.can-20-0672>
- Paul, C. L., & Clark, S. J. (1996). Cytosine Methylation: Quantitation by Automated Genomic Sequencing and GENESCAN™ Analysis. *BioTechniques*, 21(7), 126–133. <https://doi.org/10.2144/96211rr04>
- Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences*, 85(8), 2444–2448. <https://doi.org/10.1073/pnas.85.8.2444>
- Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P. G., & Fiore, P. P. D. (2010). Biological and Molecular Heterogeneity of Breast Cancers Correlates with Their Cancer Stem Cell Content. *Cell*, 140(1), 62–73. <https://doi.org/10.1016/j.cell.2009.12.007>
- Pediconi, N., Salerno, D., Lupacchini, L., Angrisani, A., Peruzzi, G., Smaele, E. D., Levrero, M., & Belloni, L. (2019). EZH2, JMJD3, and UTX epigenetically regulate hepatic plasticity inducing retro-differentiation and proliferation of liver cells. *Cell Death & Disease*, 10(7), 518. <https://doi.org/10.1038/s41419-019-1755-2>
- Peitzsch, C., Cojoc, M., Hein, L., Kurth, I., Mäbert, K., Trautmann, F., Klink, B., Schröck, E., Wirth, M. P., Krause, M., Stakhovskiy, E. A., Telegeev, G. D., Novotny, V., Toma, M., Muders, M., Baretton, G. B., Frame, F. M., Maitland, N. J., Baumann, M., & Dubrovskaya, A. (2016). An Epigenetic Reprogramming Strategy to Resensitize Radioresistant Prostate Cancer Cells. *Cancer Research*, 76(9), 2637–2651. <https://doi.org/10.1158/0008-5472.can-15-2116>
- Perekatt, A. O., Shah, P. P., Cheung, S., Jariwala, N., Wu, A., Gandhi, V., Kumar, N., Feng, Q., Patel, N., Chen, L., Joshi, S., Zhou, A., Taketo, M. M., Xing, J., White, E., Gao, N., Gatzka, M. L., & Verzi, M. (2018). SMAD4 suppresses WNT-driven de-differentiation and oncogenesis in the differentiated gut epithelium. *Cancer Research*, 78(17), canres.0043.2018. <https://doi.org/10.1158/0008-5472.can-18-0043>
- Perou, C. M., Sørlie, T., Eisen, M. B., Rijn, M. v. d., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslén, L. A., Fluge, Ø., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A.-L., Brown, P. O., & Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747–752. <https://doi.org/10.1038/35021093>
- Peter, M. E. (2009). Let-7 and miR-200 microRNAs: Guardians against pluripotency and cancer progression. *Cell Cycle*, 8(6), 843–852. <https://doi.org/10.4161/cc.8.6.7907>
- Petrova, R., & Joyner, A. L. (2014). Roles for Hedgehog signaling in adult organ homeostasis and repair. *Development*, 141(18), 3445–3457. <https://doi.org/10.1242/dev.083691>

- Pfisterer, N., Meyer-Bockenamp, F., Qu, D., Preuss, V., Rothämel, T., Geisenberger, D., Läer, K., Vennemann, B., Albers, A., Engelmann, T. A., Frieling, H., Rhein, M., & Klintschar, M. (2021). Sudden infant death syndrome revisited: serotonin transporter gene, polymorphisms and promoter methylation. *Pediatric Research*, 1–6. <https://doi.org/10.1038/s41390-021-01773-3>
- Phi, L. T. H., Sari, I. N., Yang, Y.-G., Lee, S.-H., Jun, N., Kim, K. S., Lee, Y. K., & Kwon, H. Y. (2018). Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment. *Stem Cells International*, 2018, 5416923. <https://doi.org/10.1155/2018/5416923>
- Pilar, A., Gupta, M., Laskar, S. G., & Laskar, S. (2017). Intraoperative radiotherapy: review of techniques and results. *ecancer-medicalscience*, 11, 750. <https://doi.org/10.3332/ecancer.2017.750>
- Pinamonti, M., Zanconati, F., Pinamonti, M., & Zanconati, F. (2017). Normal Breast. *Monographs in Clinical Cytology*, 24, 20–24. <https://doi.org/10.1159/000479764>
- Pinder, S. E., Duggan, C., Ellis, I. O., Cuzick, J., Forbes, J. F., Bishop, H., Fentiman, I. S., George, W. D., & Party, U. C. C. o. C. R. (D. C. I. S. (W. (2010). A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial. *British Journal of Cancer*, 103(1), 94–100. <https://doi.org/10.1038/sj.bjc.6605718>
- Pinder, S. E., & Ellis, I. O. (2003). The diagnosis and management of pre-invasive breast disease: Ductal carcinoma in situ (DCIS) and atypical ductal hyperplasia (ADH) – current definitions and classification. *Breast Cancer Research : BCR*, 5(5), 254–257. <https://doi.org/10.1186/bcr623>
- Plaks, V., Kong, N., & Werb, Z. (2015). The Cancer Stem Cell Niche: How Essential Is the Niche in Regulating Stemness of Tumor Cells? *Cell Stem Cell*, 16(3), 225–238. <https://doi.org/10.1016/j.stem.2015.02.015>
- Podergajs, N., Motaln, H., Rajčević, U., Verbovšek, U., Koršič, M., Obad, N., Espedal, H., Vittori, M., Herold-Mende, C., Miletic, H., Bjerkgvig, R., & Turnšek, T. L. (2015). Transmembrane protein CD9 is glioblastoma biomarker, relevant for maintenance of glioblastoma stem cells. *Oncotarget*, 7(1), 593–609. <https://doi.org/10.18632/oncotarget.5477>
- Pogribny, I., Koturbash, I., Tryndyak, V., Hudson, D., Stevenson, S. M., Sedelnikova, O., Bonner, W., & Kovalchuk, O. (2005). Fractionated Low-Dose Radiation Exposure Leads to Accumulation of DNA Damage and Profound Alterations in DNA and Histone Methylation in the Murine Thymus. *Molecular Cancer Research*, 3(10), 553–561. <https://doi.org/10.1158/1541-7786.mcr-05-0074>
- Pohl, S.-G., Brook, N., Agostino, M., Arfuso, F., Kumar, A. P., & Dharmarajan, A. (2017). Wnt signaling in triple-negative breast cancer. *Oncogenesis*, 6(4), e310. <https://doi.org/10.1038/oncsis.2017.14>
- Poli, V., Fagnocchi, L., & Zippo, A. (2018). Tumorigenic Cell Reprogramming and Cancer Plasticity: Interplay between Signaling, Microenvironment, and Epigenetics. *Stem Cells International*, 2018, 4598195. <https://doi.org/10.1155/2018/4598195>
- Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nature Reviews Cancer*, 9(4), 265–273. <https://doi.org/10.1038/nrc2620>
- Ponti, D., Zaffaroni, N., Capelli, C., & Daidone, M. G. (2006). Breast cancer stem cells: An overview. *European Journal of Cancer*, 42(9), 1219–1224. <https://doi.org/10.1016/j.ejca.2006.01.031>
- Postoperative radiotherapy for breast cancer: hypofractionation RCR consensus statements. (2021). <https://www.rcr.ac.uk/publication/postoperative-radiotherapy-breast-cancer-hypofractionation-rcr-consensus-statements>
- Potapov, V., & Ong, J. L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS ONE*, 12(1), e0169774. <https://doi.org/10.1371/journal.pone.0169774>
- Poucke, M. V., Boulougouris, X., Spiegeleer, B. D., Burvenich, C., Duchateau, L., & Peelman, L. J. (2017). An optimized strategy for cloning-based locus-specific bisulfite sequencing PCR. *bioRxiv*, 239566. <https://doi.org/10.1101/239566>
- Pouget, J.-P., Navarro-Teulon, I., Bardiès, M., Chouin, N., Cartron, G., Pèlegri, A., & Azria, D. (2011). Clinical radioimmunotherapy—the role of radiobiology. *Nature Reviews Clinical Oncology*, 8(12), 720–734. <https://doi.org/10.1038/nrclinonc.2011.160>
- Pradhan, S., Roberts, R. J., Bacolla, A., & Wells, R. D. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase I. EXPRESSION, PURIFICATION, AND COMPARISON OF DE NOVO AND MAINTENANCE METHYLATION*. *Journal of Biological Chemistry*, 274(46), 33002–33010. <https://doi.org/10.1074/jbc.274.46.33002>
- Prieto-Vila, M., Takahashi, R.-u., Usuba, W., Kohama, I., & Ochiya, T. (2017). Drug Resistance Driven by Cancer Stem Cells and Their Niche. *International Journal of Molecular Sciences*, 18(12), 2574. <https://doi.org/10.3390/ijms18122574>
- Proctor, E., Waghray, M., Lee, C. J., Heidt, D. G., Yalamanchili, M., Li, C., Bednar, F., & Simeone, D. M. (2013). Bmi1 Enhances Tumorigenicity and Cancer Stem Cell Function in Pancreatic Adenocarcinoma. *PLoS ONE*, 8(2), e55820. <https://doi.org/10.1371/journal.pone.0055820>
- Puhalla, S., Elmquist, W., Freyer, D., Kleinberg, L., Adkins, C., Lockman, P., McGregor, J., Muldoon, L., Nesbit, G., Peereboom, D., Smith, Q., Walker, S., & Neuwelt, E. (2015). Unsanctifying the sanctuary: challenges and opportunities with brain metastases. *Neuro-Oncology*, 17(5), 639–651. <https://doi.org/10.1093/neuonc/nov023>

Q

- Qiang, L., Wu, T., Zhang, H.-W., Lu, N., Hu, R., Wang, Y.-J., Zhao, L., Chen, F.-H., Wang, X.-T., You, Q.-D., & Guo, Q.-L. (2011). HIF-1 α is critical for hypoxia-mediated maintenance of glioblastoma stem cells by activating Notch signaling pathway. *Cell Death & Differentiation*, *19*(2), 284–294. <https://doi.org/10.1038/cdd.2011.95>
- Qiu, H., Yashiro, M., Shinto, O., Matsuzaki, T., & Hirakawa, K. (2009). DNA methyltransferase inhibitor 5-aza-CdR enhances the radiosensitivity of gastric cancer cells. *Cancer Science*, *100*(1), 181–188. <https://doi.org/10.1111/j.1349-7006.2008.01004.x>
- Qiu, P., Soder, G. J., Sanfiorenzo, V. J., Wang, L., Greene, J. R., Fritz, M. A., & Cai, X.-Y. (2003). Quantification of single nucleotide polymorphisms by automated DNA sequencing. *Biochemical and Biophysical Research Communications*, *309*(2), 331–338. <https://doi.org/10.1016/j.bbrc.2003.08.008>

R

- Raha, D., Wilson, T. R., Peng, J., Peterson, D., Yue, P., Evangelista, M., Wilson, C., Merchant, M., & Settleman, J. (2014). The Cancer Stem Cell Marker Aldehyde Dehydrogenase Is Required to Maintain a Drug-Tolerant Tumor Cell Subpopulation. *Cancer Research*, *74*(13), 3579–3590. <https://doi.org/10.1158/0008-5472.can-13-3456>
- Rainho, M. d. A., Mencialha, A. L., & Thole, A. A. (2021). Hypoxia effects on cancer stem cell phenotype in colorectal cancer: a mini-review. *Molecular Biology Reports*, *48*(11), 7527–7535. <https://doi.org/10.1007/s11033-021-06809-9>
- Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., & Beck, S. (2004). DNA Methylation Profiling of the Human Major Histocompatibility Complex: A Pilot Study for the Human Epigenome Project. *PLoS Biology*, *2*(12), e405. <https://doi.org/10.1371/journal.pbio.0020405>
- Ramakrishnan, V., Xu, B., Akers, J., Nguyen, T., Ma, J., Dhawan, S., Ning, J., Mao, Y., Hua, W., Kokkoli, E., Furnari, F., Carter, B. S., & Chen, C. C. (2020). Radiation-induced extracellular vesicle (EV) release of miR-603 promotes IGF1-mediated stem cell state in glioblastomas. *EBioMedicine*, *55*, 102736. <https://doi.org/10.1016/j.ebiom.2020.102736>
- Rao, X., Evans, J., Chae, H., Pilrose, J., Kim, S., Yan, P., Huang, R.-L., Lai, H.-C., Lin, H., Liu, Y., Miller, D., Rhee, J.-K., Huang, Y.-W., Gu, F., Gray, J. W., Huang, T.-M., & Nephew, K. P. (2013). CpG island shore methylation regulates caveolin-1 expression in breast cancer. *Oncogene*, *32*(38), 4519–4528. <https://doi.org/10.1038/onc.2012.474>
- Rappa, G., Green, T. M., Karbanová, J., Corbeil, D., & Lorico, A. (2015). Tetraspanin CD9 determines invasiveness and tumorigenicity of human breast cancer cells. *Oncotarget*, *6*(10), 7970–7991. <https://doi.org/10.18632/oncotarget.3419>
- Rauch, T., & Pfeifer, G. P. (2005). Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Laboratory Investigation*, *85*(9), 1172–1180. <https://doi.org/10.1038/labinvest.3700311>
- Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S. K., Kernstine, K. H., Riggs, A. D., & Pfeifer, G. P. (2007). Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proceedings of the National Academy of Sciences*, *104*(13), 5527–5532. <https://doi.org/10.1073/pnas.0701059104>
- Rauluseviciute, I., Drabløs, F., & Rye, M. B. (2019). DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis. *Clinical Epigenetics*, *11*(1), 193. <https://doi.org/10.1186/s13148-019-0795-x>
- Re, O. L., Douet, J., Buschbeck, M., Fusilli, C., Paziienza, V., Panebianco, C., Castracani, C. C., Mazza, T., Volti, G. L., & Vinciguerra, M. (2018). Histone variant macroH2A1 rewires carbohydrate and lipid metabolism of hepatocellular carcinoma cells towards cancer stem cells. *Epigenetics*, *13*(8), 829–845. <https://doi.org/10.1080/15592294.2018.1514239>
- Reed, K., Poulin, M. L., Yan, L., & Parissenti, A. M. (2010). Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Analytical Biochemistry*, *397*(1), 96–106. <https://doi.org/10.1016/j.ab.2009.10.021>
- Rees, W. A., Yager, T. D., Korte, J., & Hippel, P. H. V. (1993). Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry*, *32*(1), 137–144. <https://doi.org/10.1021/bi00052a019>
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, *447*(7143), 425–432. <https://doi.org/10.1038/nature05918>
- Rein, T., DePamphilis, M. L., & Zorbas, H. (1998). Identifying 5-methylcytosine and related modifications in DNA genomes. *Nucleic Acids Research*, *26*(10), 2255–2264. <https://doi.org/10.1093/nar/26.10.2255>
- Reinders, J., Vivier, C. D., Theiler, G., Chollet, D., Descombes, P., & Paszkowski, J. (2008). Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Research*, *18*(3), 469–476. <https://doi.org/10.1101/gr.7073008>

- Ren, J., Ding, L., Zhang, D., Shi, G., Xu, Q., Shen, S., Wang, Y., Wang, T., & Hou, Y. (2018). Carcinoma-associated fibroblasts promote the stemness and chemoresistance of colorectal cancer by transferring exosomal lncRNA H19. *Theranostics*, *8*(14), 3932–3948. <https://doi.org/10.7150/thno.25541>
- Renschler, M. F. (2004). The emerging role of reactive oxygen species in cancer therapy. *European Journal of Cancer*, *40*(13), 1934–1940. <https://doi.org/10.1016/j.ejca.2004.02.031>
- Rheinbay, E., Suvà, M. L., Gillespie, S. M., Wakimoto, H., Patel, A. P., Shahid, M., Oksuz, O., Rabkin, S. D., Martuza, R. L., Rivera, M. N., Louis, D. N., Kasif, S., Chi, A. S., & Bernstein, B. E. (2013). An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell Reports*, *3*(5), 1567–1579. <https://doi.org/10.1016/j.celrep.2013.04.021>
- Rhim, A. D., Mirek, E. T., Aiello, N. M., Maitra, A., Bailey, J. M., McAllister, F., Reichert, M., Beatty, G. L., Rustgi, A. K., Vonderheide, R. H., Leach, S. D., & Stanger, B. Z. (2012). EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell*, *148*(1-2), 349–361. <https://doi.org/10.1016/j.cell.2011.11.025>
- Rich, J. N. (2016). Cancer stem cells: understanding tumor hierarchy and heterogeneity. *Medicine*, *95*(Suppl 1), S2–S7. <https://doi.org/10.1097/md.00000000000004764>
- Rios, A. C., Fu, N. Y., Lindeman, G. J., & Visvader, J. E. (2014). In situ identification of bipotent stem cells in the mammary gland. *Nature*, *506*(7488), 322–327. <https://doi.org/10.1038/nature12948>
- Robinson, J. T., Thorvaldsdóttir, H., Turner, D., & Mesirov, J. P. (2020). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *bioRxiv*, 2020.05.03.075499. <https://doi.org/10.1101/2020.05.03.075499>
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Research*, *77*(21), e31–e34. <https://doi.org/10.1158/0008-5472.can-17-0337>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature biotechnology*, *29*(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rodda, D. J., Chew, J.-L., Lim, L.-H., Loh, Y.-H., Wang, B., Ng, H.-H., & Robson, P. (2005). Transcriptional Regulation of Nanog by OCT4 and SOX2*. *Journal of Biological Chemistry*, *280*(26), 24731–24737. <https://doi.org/10.1074/jbc.m502573200>
- Rodrigues, C. F. D., Serrano, E., Patrício, M. I., Val, M. M., Albuquerque, P., Fonseca, J., Gomes, C. M. F., Abrunhosa, A. J., Paiva, A., Carvalho, L., Botelho, M. F., Almeida, L., Carreira, I. M., & Alpoim, M. C. (2018). Stroma-derived IL-6, G-CSF and Activin-A mediated dedifferentiation of lung carcinoma cells into cancer stem cells. *Scientific Reports*, *8*(1), 11573. <https://doi.org/10.1038/s41598-018-29947-w>
- Rodrigues, M., Kosaric, N., Bonham, C. A., & Gurtner, G. C. (2019). Wound Healing: A Cellular Perspective. *Physiological Reviews*, *99*(1), 665–706. <https://doi.org/10.1152/physrev.00067.2017>
- Rohde, C., Zhang, Y., Reinhardt, R., & Jeltsch, A. (2010). BISMA - Fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*, *11*(1), 230–230. <https://doi.org/10.1186/1471-2105-11-230>
- Roscigno, G., Quintavalle, C., Donnarumma, E., Puoti, I., Diaz-Lagares, A., Iaboni, M., Fiore, D., Russo, V., Todaro, M., Romano, G., Thomas, R., Cortino, G., Gaggianesi, M., Esteller, M., Croce, C. M., & Condorelli, G. (2015). MiR-221 promotes stemness of breast cancer cells by targeting DNMT3b. *Oncotarget*, *7*(1), 580–592. <https://doi.org/10.18632/oncotarget.5979>
- Rosen, E. M., Fan, S., Pestell, R. G., & Goldberg, I. D. (2003). BRCA1 gene in breast cancer. *Journal of Cellular Physiology*, *196*(1), 19–41. <https://doi.org/10.1002/jcp.10257>
- Roux, K. H. (2009). Optimization and Troubleshooting in PCR. *Cold Spring Harbor Protocols*, 2009(4), pdb.ip66. <https://doi.org/10.1101/pdb.ip66>
- Roy, H. B.-L., Vennin, C., Brocqueville, G., Spruyt, N., Adriaenssens, E., & Bourette, R. P. (2015). Enrichment of Human Stem-Like Prostate Cells with s-SHIP Promoter Activity Uncovers a Role in Stemness for the Long Noncoding RNA H19. *Stem Cells and Development*, *24*(10), 1252–1262. <https://doi.org/10.1089/scd.2014.0386>
- Roy, P. G., & Thompson, A. M. (2006). Cyclin D1 and breast cancer. *The Breast*, *15*(6), 718–727. <https://doi.org/10.1016/j.breast.2006.02.005>
- Roy, R., Chun, J., & Powell, S. N. (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews Cancer*, *12*(1), 68–78. <https://doi.org/10.1038/nrc3181>
- Rubatino, F., Carobin, N., Freitas, M., Oliveira, V. d., Pietra, R., Oliveira, P., Bosco, A., & Jehee, F. (2015). Manipulation of primer affinity improves high-resolution melting accuracy for imprinted genes. *Genetics and Molecular Research*, *14*(3), 7864–7872. <https://doi.org/10.4238/2015.july.14.12>

S

- Safa, A. R., Saadatzaadeh, M. R., Cohen-Gadol, A. A., Pollok, K. E., & Bijangi-Vishehsaraei, K. (2015). Glioblastoma stem cells (GSCs) epigenetic plasticity and interconversion between differentiated non-GSCs and GSCs. *Genes & Diseases*, 2(2), 152–163. <https://doi.org/10.1016/j.gendis.2015.02.001>
- Saha, S., Mukherjee, S., Khan, P., Kajal, K., Mazumdar, M., Manna, A., Mukherjee, S., De, S., Jana, D., Sarkar, D. K., & Das, T. (2016). Aspirin Suppresses the Acquisition of Chemoresistance in Breast Cancer by Disrupting an NFκB-IL6 Signaling Axis Responsible for the Generation of Cancer Stem Cells. *Cancer Research*, 76(7), 2000–2012. <https://doi.org/10.1158/0008-5472.can-15-1360>
- Saijo, H., Hirohashi, Y., Torigoe, T., Horibe, R., Takaya, A., Murai, A., Kubo, T., Kajiwara, T., Tanaka, T., Shionoya, Y., Yamamoto, E., Maruyama, R., Nakatsugawa, M., Kanaseki, T., Tsukahara, T., Tamura, Y., Sasaki, Y., Tokino, T., Suzuki, H., . . . Sato, N. (2016). Plasticity of lung cancer stem-like cells is regulated by the transcription factor HOXA5 that is induced by oxidative stress. *Oncotarget*, 7(31), 50043–50056. <https://doi.org/10.18632/oncotarget.10571>
- Sakaki-Yumoto, M., Katsuno, Y., & Derynck, R. (2013). TGF-β family signaling in stem cells. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1830(2), 2280–2296. <https://doi.org/10.1016/j.bbagen.2012.08.008>
- Saksornchai, K., Jaruthien, T., Nantavithya, C., Shotelersuk, K., & Rojpornpradit, P. (2021). Long-term results of hypofractionation with concomitant boost in patients with early breast cancer: A prospective study. *PLoS ONE*, 16(10), e0258186. <https://doi.org/10.1371/journal.pone.0258186>
- Salnikov, A. V., Iversen, V. V., Koisti, M., Sundberg, C., Johansson, L., Stuhr, L. B., Sjöquist, M., Ahlström, H., Reed, R. K., & Rubin, K. (2003). Lowering of tumor interstitial fluid pressure specifically augments efficacy of chemotherapy. *The FASEB Journal*, 17(12), 1756–1758. <https://doi.org/10.1096/fj.02-1201fje>
- Sanaa, E. M. (2022). cancer cell de-differentiation: plasticity- driven stratagem for tumor metastasis and recurrence. *Current Stem Cell Research & Therapy*, 17. <https://doi.org/10.2174/1574888x17666220608101852>
- Sandri, C., Caccavari, F., Valdembrì, D., Camillo, C., Veltel, S., Santambrogio, M., Lanzetti, L., Bussolino, F., Ivaska, J., & Serini, G. (2012). The R-Ras/RIN2/Rab5 complex controls endothelial cell adhesion and morphogenesis via active integrin endocytosis and Rac signaling. *Cell Research*, 22(10), 1479–1501. <https://doi.org/10.1038/cr.2012.110>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- SanMiguel, J. M., & Bartolomei, M. S. (2018). DNA methylation dynamics of genomic imprinting in mouse development. *Biology of Reproduction*, 99(1), 252–262. <https://doi.org/10.1093/biolre/iy036>
- Sansregret, L., Vanhaesebroeck, B., & Swanton, C. (2018). Determinants and clinical implications of chromosomal instability in cancer. *Nature Reviews Clinical Oncology*, 15(3), 139–150. <https://doi.org/10.1038/nrclinonc.2017.198>
- Santenard, A., & Torres-Padilla, M.-E. (2009). Epigenetic reprogramming in mammalian reproduction: Contribution from histone variants. *Epigenetics*, 4(2), 80–84. <https://doi.org/10.4161/epi.4.2.7838>
- Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, 103(5), 1412–1417. <https://doi.org/10.1073/pnas.0510310103>
- Scheel, C., & Weinberg, R. A. (2012). Cancer stem cells and epithelial-mesenchymal transition: Concepts and molecular links. *Seminars in Cancer Biology*, 22(5-6), 396–403. <https://doi.org/10.1016/j.semcancer.2012.04.001>
- Schiele, M. A., Zwanzger, P., Schwarte, K., Arolt, V., Baune, B. T., & Domschke, K. (2021). Serotonin Transporter Gene Promoter Hypomethylation as a Predictor of Antidepressant Treatment Response in Major Depression: A Replication Study. *International Journal of Neuropsychopharmacology*, 24(3), 191–199. <https://doi.org/10.1093/ijnp/pyaa081>
- Schillert, A., Trumpp, A., & Sprick, M. R. (2013). Label retaining cells in cancer - The dormant root of evil? *Cancer Letters*, 341(1), 73–79. <https://doi.org/10.1016/j.canlet.2013.04.019>
- Schmitt, F., Oakeley, E. J., & Jost, J. P. (1997). Antibiotics Induce Genome-wide Hypermethylation in Cultured *Nicotiana tabacum* Plants*. *Journal of Biological Chemistry*, 272(3), 1534–1540. <https://doi.org/10.1074/jbc.272.3.1534>
- Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, 517(7534), 321–326. <https://doi.org/10.1038/nature14192>
- Schulz, A., Meyer, F., Dubrovskaya, A., & Borgmann, K. (2019). Cancer Stem Cells and Radioresistance: DNA Repair and Beyond. *Cancers*, 11(6), 862. <https://doi.org/10.3390/cancers11060862>
- Schwab, L. P., Peacock, D. L., Majumdar, D., Ingels, J. F., Jensen, L. C., Smith, K. D., Cushing, R. C., & Seagroves, T. N. (2012). Hypoxia-inducible factor 1α promotes primary tumor growth and tumor-initiating cell activity in breast cancer. *Breast Cancer Research : BCR*, 14(1), R6–R6. <https://doi.org/10.1186/bcr3087>
- Schwitala, S., Fingerle, A. A., Cammareri, P., Nebelsiek, T., Göktuna, S. I., Ziegler, P. K., Canli, O., Heijmans, J., Huels, D. J., Moreaux, G., Rupec, R. A., Gerhard, M., Schmid, R., Barker, N., Clevers, H., Lang, R., Neumann, J., Kirchner, T.,

- Taketo, M. M., ... Greten, F. R. (2013). Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. *Cell*, 152(1-2), 25–38. <https://doi.org/10.1016/j.cell.2012.12.012>
- Sehn, J. K. (2015). Clinical Genomics. *Section II: Bioinformatics*, 129–150. <https://doi.org/10.1016/b978-0-12-404748-8.00009-5>
- Sekar, D., Krishnan, R., Panagal, M., Sivakumar, P., Gopinath, V., & Basam, V. (2016). Deciphering the role of microRNA 21 in cancer stem cells (CSCs). *Genes & Diseases*, 3(4), 277–281. <https://doi.org/10.1016/j.gendis.2016.05.002>
- Semenza, G. L. (2016a). Dynamic regulation of stem cell specification and maintenance by hypoxia-inducible factors. *Molecular Aspects of Medicine*, 47-48, 15–23. <https://doi.org/10.1016/j.mam.2015.09.004>
- Semenza, G. L. (2016b). Hypoxia-inducible factors: coupling glucose metabolism and redox regulation with induction of the breast cancer stem cell phenotype. *The EMBO journal*, 36(3), 252–259. <https://doi.org/10.15252/emboj.201695204>
- Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M.-L., Wu, L., Lindeman, G. J., & Visvader, J. E. (2006). Generation of a functional mammary gland from a single stem cell. *Nature*, 439(7072), 84–88. <https://doi.org/10.1038/nature04372>
- Shapiro, R., Braverman, B., Louis, J. B., & Servis, R. E. (1973). Nucleic Acid Reactivity and Conformation II. REACTION OF CYTOSINE AND URACIL WITH SODIUM BISULFITE. *Journal of Biological Chemistry*, 248(11), 4060–4064. [https://doi.org/10.1016/s0021-9258\(19\)43838-6](https://doi.org/10.1016/s0021-9258(19)43838-6)
- Shapiro, R., Servis, R. E., & Welcher, M. (1970). Reactions of Uracil and Cytosine Derivatives with Sodium Bisulfite. *Journal of the American Chemical Society*, 92(2), 422–424. <https://doi.org/10.1021/ja00705a626>
- Sharma, P., Hu-Lieskovan, S., Wargo, J. A., & Ribas, A. (2017). Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell*, 168(4), 707–723. <https://doi.org/10.1016/j.cell.2017.01.017>
- Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I. A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C. J., & Stingl, J. (2012). Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research : BCR*, 14(5), R134–R134. <https://doi.org/10.1186/bcr3334>
- Shelly, W., Draper, M. W., Krishnan, V., Wong, M., & Jaffe, R. B. (2008). Selective Estrogen Receptor Modulators; An Update on Recent Clinical Findings. *Obstetrical & Gynecological Survey*, 63(3), 163–181. <https://doi.org/10.1097/ogx.0b013e31816400d7>
- Shen, L., Guo, Y., Chen, X., Ahmed, S., & Issa, J.-P. J. (2007). Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. *BioTechniques*, 42(1), 48–58. <https://doi.org/10.2144/000112312>
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H.-L., Zhang, K., & Zhang, Y. (2013). Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell*, 153(3), 692–706. <https://doi.org/10.1016/j.cell.2013.04.002>
- Shen, M., Dong, C., Ruan, X., Yan, W., Cao, M., Pizzo, D., Wu, X., Yang, L., Liu, L., Ren, X., & Wang, S. E. (2019). Chemotherapy-Induced Extracellular Vesicle miRNAs Promote Breast Cancer Stemness by Targeting ONECUT2. *Cancer Research*, 79(14), 3608–3621. <https://doi.org/10.1158/0008-5472.can-18-4055>
- Shi, J., Whyte, W. A., Zepeda-Mendoza, C. J., Milazzo, J. P., Shen, C., Roe, J.-S., Minder, J. L., Mercan, F., Wang, E., Eckersley-Maslin, M. A., Campbell, A. E., Kawaoka, S., Shareef, S., Zhu, Z., Kendall, J., Muhar, M., Haslinger, C., Yu, M., Roeder, R. G., ... Vakoc, C. R. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes & Development*, 27(24), 2648–2662. <https://doi.org/10.1101/gad.232710.113>
- Shimono, Y., Mukohyama, J., Nakamura, S.-i., & Minami, H. (2015). MicroRNA Regulation of Human Breast Cancer Stem Cells. *Journal of Clinical Medicine*, 5(1), 2. <https://doi.org/10.3390/jcm5010002>
- Shimono, Y., Zabala, M., Cho, R. W., Lobo, N., Dalerba, P., Qian, D., Diehn, M., Liu, H., Panula, S. P., Chiao, E., Dirbas, F. M., Somlo, G., Pera, R. A. R., Lao, K., & Clarke, M. F. (2009). Downregulation of miRNA-200c Links Breast Cancer Stem Cells with Normal Stem Cells. *Cell*, 138(3), 592–603. <https://doi.org/10.1016/j.cell.2009.07.011>
- Shinde, D., Lai, Y., Sun, F., & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)_n and (A/T)_n microsatellites. *Nucleic Acids Research*, 31(3), 974–980. <https://doi.org/10.1093/nar/gkg178>
- Shipitsin, M., & Polyak, K. (2008). The cancer stem cell hypothesis: in search of definitions, markers, and relevance. *Laboratory Investigation*, 88(5), 459–463. <https://doi.org/10.1038/labinvest.2008.14>
- Shipony, Z., Mukamel, Z., Cohen, N. M., Landan, G., Chomsky, E., Zeliger, S. R., Fried, Y. C., Ainsbinder, E., Friedman, N., & Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature*, 513(7516), 115–119. <https://doi.org/10.1038/nature13458>
- Shuang, Z.-Y., Wu, W.-C., Xu, J., Lin, G., Liu, Y.-C., Lao, X.-M., Zheng, L., & Li, S. (2014). Transforming growth factor- β -induced epithelial-mesenchymal transition generates ALDH-positive cells with stem cell properties in cholangiocarcinoma. *Cancer Letters*, 354(2), 320–328. <https://doi.org/10.1016/j.canlet.2014.08.030>

- Silva, J. L. d., Nunes, N. C. C., Izetti, P., Mesquita, G. G. d., & Melo, A. C. d. (2019). Triple Negative Breast Cancer: a thorough review of biomarkers. *Critical Reviews in Oncology/Hematology*, *145*, 102855. <https://doi.org/10.1016/j.critrevonc.2019.102855>
- Silva-Diz, V. d., Lorenzo-Sanz, L., Bernat-Peguera, A., Lopez-Cerda, M., & Muñoz, P. (2018). Cancer cell plasticity: Impact on tumor progression and therapy response. *Seminars in Cancer Biology*, *53*, 48–58. <https://doi.org/10.1016/j.semcancer.2018.08.009>
- Singer-Sam, J., Grant, M., LeBon, J. M., Okuyama, K., Chapman, V., Monk, M., & Riggs, A. D. (1990). Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pcg-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Molecular and Cellular Biology*, *10*(9), 4987–4989. <https://doi.org/10.1128/mcb.10.9.4987>
- Singh, S. K., Hawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., Henkelman, R. M., Cusimano, M. D., & Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature*, *432*(7015), 396–401. <https://doi.org/10.1038/nature03128>
- Skvortsov, S., Debbage, P., Lukas, P., & Skvortsova, I. (2015). Crosstalk between DNA repair and cancer stem cell (CSC) associated intracellular pathways. *Seminars in Cancer Biology*, *31*, 36–42. <https://doi.org/10.1016/j.semcancer.2014.06.002>
- Skvortsova, I., Debbage, P., Kumar, V., & Skvortsov, S. (2015). Radiation resistance: Cancer stem cells (CSCs) and their enigmatic pro-survival signaling. *Seminars in Cancer Biology*, *35*, 39–44. <https://doi.org/10.1016/j.semcancer.2015.09.009>
- Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., & McGuire, W. L. (1987). Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. *Science*, *235*(4785), 177–182. <https://doi.org/10.1126/science.3798106>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59. <https://doi.org/10.1002/cpmb.59>
- Slatko, B. E., Kieleczawa, J., Ju, J., Gardner, A. F., Hendrickson, C. L., & Ausubel, F. M. (2011). "First Generation" Automated DNA Sequencing Technology. *Current Protocols in Molecular Biology*, *96*(1), 7.2.1–7.2.28. <https://doi.org/10.1002/0471142727.mb0702s96>
- Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S. L., Digles, D., Ehrhart, F., Giesbertz, P., Kalafati, M., Martens, M., Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L. M. T., . . . Willighagen, E. L. (2018). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*, *46*(Database issue), D661–D667. <https://doi.org/10.1093/nar/gkx1064>
- Smigiel, J. M., Parameswaran, N., & Jackson, M. W. (2017). Potent EMT and CSC Phenotypes Are Induced By Oncostatin-M in Pancreatic Cancer. *Molecular Cancer Research*, *15*(4), 478–488. <https://doi.org/10.1158/1541-7786.mcr-16-0337>
- Smigiel, J. M., Parameswaran, N., & Jackson, M. W. (2018). Targeting Pancreatic Cancer Cell Plasticity: The Latest in Therapeutics. *Cancers*, *10*(1), 14. <https://doi.org/10.3390/cancers10010014>
- Smith, L. M., Fung, S., Hunkapiller, M. W., Hunkapiller, T. J., & Hood, L. E. (1985). The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Research*, *13*(7), 2399–2412. <https://doi.org/10.1093/nar/13.7.2399>
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, *321*(6071), 674–679. <https://doi.org/10.1038/321674a0>
- Smith, Z. D., Gu, H., Bock, C., Gnirke, A., & Meissner, A. (2009). High-throughput bisulfite sequencing in mammalian genomes. *Methods*, *48*(3), 226–232. <https://doi.org/10.1016/j.ymeth.2009.05.003>
- Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. <https://doi.org/10.1038/nrg3354>
- Soeda, A., Park, M., Lee, D., Mintz, A., Androutsellis-Theotokis, A., McKay, R. D., Engh, J., Iwama, T., Kunisada, T., Kassam, A. B., Pollack, I. F., & Park, D. M. (2009). Hypoxia promotes expansion of the CD133-positive glioma stem cells through activation of HIF-1 α . *Oncogene*, *28*(45), 3949–3959. <https://doi.org/10.1038/onc.2009.252>
- Song, Q., Decato, B., Hong, E. E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., & Smith, A. D. (2013a). A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. *PLoS ONE*, *8*(12), e81148. <https://doi.org/10.1371/journal.pone.0081148>
- Song, S. J., Poliseno, L., Song, M. S., Ala, U., Webster, K., Ng, C., Beringer, G., Brikbak, N. J., Yuan, X., Cantley, L. C., Richardson, A. L., & Pandolfi, P. P. (2013b). MicroRNA-Antagonism Regulates Breast Cancer Stemness and Metastasis via TET-Family-Dependent Chromatin Remodeling. *Cell*, *154*(2), 311–324. <https://doi.org/10.1016/j.cell.2013.06.026>

- Song, X.-y., Li, B.-y., Zhou, E.-x., & Wu, F.-x. (2016). The clinicopathological significance of RUNX3 hypermethylation and mRNA expression in human breast cancer, a meta-analysis. *OncoTargets and therapy*, 9, 5339–5347. <https://doi.org/10.2147/ott.s77828>
- Song, Y., Wang, Y., Xu, L., Ma, J., Chen, E., Zang, R., Jia, W., Tao, X., & Hu, L. (2015). A genetic variant in CHRN3–CHRNA6 increases risk of esophageal squamous cell carcinoma in Chinese populations. *Carcinogenesis*, 36(5), 538–542. <https://doi.org/10.1093/carcin/bgv019>
- Soozangar, N., Sadeghi, M. R., Jeddi, F., Somi, M. H., Shirmohamadi, M., & Samadi, N. (2018). Comparison of genome-wide analysis techniques to DNA methylation analysis in human cancer. *Journal of Cellular Physiology*, 233(5), 3968–3981. <https://doi.org/10.1002/jcp.26176>
- Sørli, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., Rijn, M. v. d., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E., & Børresen-Dale, A.-L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98(19), 10869–10874. <https://doi.org/10.1073/pnas.191367098>
- Speers, C., & Pierce, L. J. (2016). Postoperative Radiotherapy After Breast-Conserving Surgery for Early-Stage Breast Cancer: A Review. *JAMA Oncology*, 2(8), 1075. <https://doi.org/10.1001/jamaoncol.2015.5805>
- Sporn, M. (1996). The war on cancer. *The Lancet*, 347(9012), 1377–1381. [https://doi.org/10.1016/s0140-6736\(96\)91015-6](https://doi.org/10.1016/s0140-6736(96)91015-6)
- Spring, L. M., Bar, Y., & Isakoff, S. J. (2022). The Evolving Role of Neoadjuvant Therapy for Operable Breast Cancer. *Journal of the National Comprehensive Cancer Network*, 20(6), 723–734. <https://doi.org/10.6004/jnccn.2022.7016>
- Steeg, P. S., & Zhou, Q. (1998). Cyclins and breast cancer. *Breast Cancer Research and Treatment*, 52(1-3), 17–28. <https://doi.org/10.1023/a:1006102916060>
- Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M.-L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., . . . Campbell, P. J. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell*, 144(1), 27–40. <https://doi.org/10.1016/j.cell.2010.11.055>
- Stirzaker, C., Millar, D. S., Paul, C. L., Warnecke, P. M., Harrison, J., Vincent, P. C., Frommer, M., & Clark, S. J. (1997). Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer research*, 57(11), 2229–37.
- Stricker, S., & Pollard, S. (2014). Reprogramming cancer cells to pluripotency. *Epigenetics*, 9(6), 798–802. <https://doi.org/10.4161/epi.28600>
- Stuelten, C. H., Mertins, S. D., Busch, J. I., Gowens, M., Scudiero, D. A., Burkett, M. W., Hite, K. M., Alley, M., Hollingshead, M., Shoemaker, R. H., & Niederhuber, J. E. (2010). Complex Display of Putative Tumor Stem Cell Markers in the NCI60 Tumor Cell Line Panel. *STEM CELLS*, 28(4), 649–660. <https://doi.org/10.1002/stem.324>
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., & Tajima, S. (2004). DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *Journal of Biological Chemistry*, 279(26), 27816–27823. <https://doi.org/10.1074/jbc.m400181200>
- Sun, L., Mathews, L. A., Cabarcas, S. M., Zhang, X., Yang, A., Zhang, Y., Young, M. R., Klarmann, K. D., Keller, J. R., & Farrar, W. L. (2013). Epigenetic Regulation of SOX9 by the NF-κB Signaling Pathway in Pancreatic Cancer Stem Cells. *STEM CELLS*, 31(8), 1454–1466. <https://doi.org/10.1002/stem.1394>
- Sun, X., Bao, J., You, Z., Chen, X., & Cui, J. (2016). Modeling of signaling crosstalk-mediated drug resistance and its implications on drug combination. *Oncotarget*, 7(39), 63995–64006. <https://doi.org/10.18632/oncotarget.11745>
- Sun, Z., & Yan, B. (2020). Multiple roles and regulatory mechanisms of the transcription factor GATA6 in human cancers. *Clinical Genetics*, 97(1), 64–72. <https://doi.org/10.1111/cge.13630>
- Sun, Z., Wang, L., Dong, L., & Wang, X. (2018). Emerging role of exosome signalling in maintaining cancer stem cell dynamic equilibrium. *Journal of Cellular and Molecular Medicine*, 22(8), 3719–3728. <https://doi.org/10.1111/jcmm.13676>
- Sun, Z., Wang, L., Zhou, Y., Dong, L., Ma, W., Lv, L., Zhang, J., & Wang, X. (2020). Glioblastoma Stem Cell-Derived Exosomes Enhance Stemness and Tumorigenicity of Glioma Cells by Transferring Notch1 Protein. *Cellular and Molecular Neurobiology*, 40(5), 767–784. <https://doi.org/10.1007/s10571-019-00771-8>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Sutherland, K. D., & Visvader, J. E. (2015). Cellular Mechanisms Underlying Intertumoral Heterogeneity. *Trends in Cancer*, 1(1), 15–23. <https://doi.org/10.1016/j.trecan.2015.07.003>
- Sutton, L. P., Jeffreys, S. A., Phillips, J. L., Taberlay, P. C., Holloway, A. F., Ambrose, M., Joo, J.-H. E., Young, A., Berry, R., Skala, M., & Brettingham-Moore, K. H. (2019). DNA methylation changes following DNA damage in prostate cancer cells. *Epigenetics*, 14(10), 1–14. <https://doi.org/10.1080/15592294.2019.1629231>

- Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., ... Bernstein, B. E. (2014). Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. *Cell*, *157*(3), 580–594. <https://doi.org/10.1016/j.cell.2014.02.030>
- Suvà, M. L., Riggi, N., & Bernstein, B. E. (2013). Epigenetic Reprogramming in Cancer. *Science*, *339*(6127), 1567–1570. <https://doi.org/10.1126/science.1230184>
- Suvà, M.-L., Riggi, N., Janiszewska, M., Radovanovic, I., Provero, P., Stehle, J.-C., Baumer, K., Bitoux, M.-A. L., Marino, D., Cironi, L., Marquez, V. E., Clément, V., & Stamenkovic, I. (2009). EZH2 Is Essential for Glioblastoma Cancer Stem Cell Maintenance. *Cancer Research*, *69*(24), 9211–9218. <https://doi.org/10.1158/0008-5472.can-09-1622>
- Suzuki, H., Itoh, F., Toyota, M., Kikuchi, T., Kakiuchi, H., Hinoda, Y., & Imai, K. (2000). Quantitative DNA methylation analysis by fluorescent polymerase chain reaction single-strand conformation polymorphism using an automated DNA sequencer. *ELECTROPHORESIS*, *21*(5), 904–908. [https://doi.org/10.1002/\(sici\)1522-2683\(20000301\)21:5<904::aid-elps904>3.0.co;2-4](https://doi.org/10.1002/(sici)1522-2683(20000301)21:5<904::aid-elps904>3.0.co;2-4)
- Szatmári, T., Kis, D., Bogdándi, E. N., Benedek, A., Bright, S., Bowler, D., Persa, E., Kis, E., Balogh, A., Naszályi, L. N., Kadhim, M., Sáfrány, G., & Lumniczky, K. (2017). Extracellular Vesicles Mediate Radiation-Induced Systemic Bystander Signals in the Bone Marrow and Spleen. *Frontiers in Immunology*, *8*, 347. <https://doi.org/10.3389/fimmu.2017.00347>
- Szymanowska-Narloch, A., Jassem, E., Skrzypski, M., Muley, T., Meister, M., Dienemann, H., Taron, M., Rosell, R., Rzepko, R., Jarzab, M., Marjański, T., Pawłowski, R., Rzyman, W., & Jassem, J. (2013). Molecular profiles of non-small cell lung cancers in cigarette smoking and never-smoking patients. *Advances in Medical Sciences*, *58*(2), 196–206. <https://doi.org/10.2478/ams-2013-0025>

T

- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. <https://doi.org/10.1016/j.cell.2007.11.019>
- Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Talati, P. G., Gu, L., Ellsworth, E. M., Gironde, M. A., Tretotola, M., Hoang, D. T., Leiby, B., Dagvadorj, A., McCue, P. A., Lallas, C. D., Trabulsi, E. J., Gomella, L., Aplin, A. E., Languino, L., Fatatis, A., Rui, H., & Nevalainen, M. T. (2015). Jak2-Stat5a/b Signaling Induces Epithelial-to-Mesenchymal Transition and Stem-Like Cell Properties in Prostate Cancer. *The American Journal of Pathology*, *185*(9), 2505–2522. <https://doi.org/10.1016/j.ajpath.2015.04.026>
- Tan, H.-W., Leung, C. O.-N., Chan, K. K.-S., Ho, D. W.-H., Leung, M.-S., Wong, C.-M., Ng, I. O.-L., & Lo, R. C.-L. (2019). Deregulated GATA6 modulates stem cell-like properties and metabolic phenotype in hepatocellular carcinoma. *International Journal of Cancer*, *145*(7), 1860–1873. <https://doi.org/10.1002/ijc.32248>
- Tanaka, H., Nakamura, M., Kameda, C., & Anticancer ... , K. M. (2009). The Hedgehog signaling pathway plays an essential role in maintaining the CD44+ CD24-/low subpopulation and the side population of breast cancer cells.
- Tanderup, K., Ménard, C., Polgar, C., Lindegaard, J. C., Kirisits, C., & Pötter, R. (2017). Advancements in brachytherapy. *Advanced Drug Delivery Reviews*, *109*, 15–25. <https://doi.org/10.1016/j.addr.2016.09.002>
- Tanei, T., Morimoto, K., Shimazu, K., Kim, S., Tanji, Y., Taguchi, T., Tamaki, Y., & Noguchi, S. (2009). Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*, *15*(12), 4234–41. <https://doi.org/10.1158/1078-0432.ccr-08-1479>
- Tang, X., Li, X., Li, Z., Liu, Y., Yao, L., Song, S., Yang, H., & Li, C. (2016). Downregulation of CXCR7 inhibits proliferative capacity and stem cell-like properties in breast cancer stem cells. *Tumor Biology*, *37*(10), 13425–13433. <https://doi.org/10.1007/s13277-016-5180-1>
- Thankamony, A. P., Saxena, K., Murali, R., Jolly, M. K., & Nair, R. (2020). Cancer Stem Cell Plasticity – A Deadly Deal. *Frontiers in Molecular Biosciences*, *7*, 79. <https://doi.org/10.3389/fmolb.2020.00079>
- Thé, H. d. (2018). Differentiation therapy revisited. *Nature Reviews Cancer*, *18*(2), 117–127. <https://doi.org/10.1038/nrc.2017.103>
- The French National Cancer Institute (INCa). (2022). <https://www.e-cancer.fr/>
- Theunissen, T. W., Oosten, A. L. v., Castelo-Branco, G., Hall, J., Smith, A., & Silva, J. C. (2011). Nanog Overcomes Reprogramming Barriers and Induces Pluripotency in Minimal Conditions. *Current Biology*, *21*(1), 65–71. <https://doi.org/10.1016/j.cub.2010.11.074>
- Thienpont, B., Dyck, L. V., & Lambrechts, D. (2016). Tumors smother their epigenome. *Molecular & Cellular Oncology*, *3*(6), e1240549. <https://doi.org/10.1080/23723556.2016.1240549>

- Thomas, P., Pranatharthi, A., Ross, C., & Srivastava, S. (2019). RhoC: a fascinating journey from a cytoskeletal organizer to a Cancer stem cell therapeutic target. *Journal of Experimental & Clinical Cancer Research : CR*, *38*(1), 328. <https://doi.org/10.1186/s13046-019-1327-4>
- Thomassin, H., Kress, C., & Grange, T. (2004). MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. *Nucleic Acids Research*, *32*(21), e168–e168. <https://doi.org/10.1093/nar/gnh166>
- Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A., & Ramanathan, S. (2011). Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell*, *145*(6), 875–889. <https://doi.org/10.1016/j.cell.2011.05.017>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Tian, J., Lee, S. O., Liang, L., Luo, J., Huang, C.-K., Li, L., Niu, Y., & Chang, C. (2012). Targeting the Unique Methylation Pattern of Androgen Receptor (AR) Promoter in Prostate Stem/Progenitor Cells with 5-Aza-2'-deoxycytidine (5-AZA) Leads to Suppressed Prostate Tumorigenesis. *Journal of Biological Chemistry*, *287*(47), 39954–39966. <https://doi.org/10.1074/jbc.m112.395574>
- Toh, T. B., Lim, J. J., & Chow, E. K.-H. (2017). Epigenetics in cancer stem cells. *Molecular Cancer*, *16*(1), 29. <https://doi.org/10.1186/s12943-017-0596-9>
- Topalovic, V., Schwirtlich, M., Stevanovic, M., & Mojsin, M. (2017). Histone modifications on the promoters of human OCT4 and NANOG genes at the onset of neural differentiation of NT2/D1 cells. *Biochemistry (Moscow)*, *82*(6), 715–722. <https://doi.org/10.1134/s0006297917060086>
- Torres, L., Ribeiro, F. R., Pandis, N., Andersen, J. A., Heim, S., & Teixeira, M. R. (2007). Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast Cancer Research and Treatment*, *102*(2), 143–155. <https://doi.org/10.1007/s10549-006-9317-6>
- Tost, J. (2010). DNA Methylation: An Introduction to the Biology and the Disease-Associated Changes of a Promising Biomarker. *Molecular Biotechnology*, *44*(1), 71–81. <https://doi.org/10.1007/s12033-009-9216-2>
- Tower, H., Ruppert, M., & Britt, K. (2019). The Immune Microenvironment of Breast Cancer Progression. *Cancers*, *11*(9), 1375. <https://doi.org/10.3390/cancers11091375>
- Toyota, M., Ho, C., Ahuja, N., Jair, K. W., Li, Q., Ohe-Toyota, M., Baylin, S. B., & Issa, J. P. (1999). Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer research*, *59*(10), 2307–12.
- Tsai, H.-C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F. V., Shin, J. J., Harbom, K. M., Beaty, R., Pappou, E., Harris, J., Yen, R.-W. C., Ahuja, N., Brock, M. V., Stearns, V., Feller-Kopman, D., Yarmus, L. B., Lin, Y.-C., Welm, A. L., ... Zahnow, C. A. (2012). Transient Low Doses of DNA-Demethylating Agents Exert Durable Antitumor Effects on Hematological and Epithelial Tumor Cells. *Cancer Cell*, *21*(3), 430–446. <https://doi.org/10.1016/j.ccr.2011.12.029>
- Turashvili, G., & Brogi, E. (2017). Tumor Heterogeneity in Breast Cancer. *Frontiers in Medicine*, *4*, 227. <https://doi.org/10.3389/fmed.2017.00227>
- Turinetto, V., & Giachino, C. (2015). Histone variants as emerging regulators of embryonic stem cell identity. *Epigenetics*, *10*(7), 563–573. <https://doi.org/10.1080/15592294.2015.1053682>
- Tusnády, G. E., Simon, I., Váradi, A., & Arányi, T. (2005). BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Research*, *33*(1), e9–e9. <https://doi.org/10.1093/nar/gni012>

U

- Uhlmann, K., Brinckmann, A., Toliat, M. R., Ritter, H., & Nürnberg, P. (2002). Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *ELECTROPHORESIS*, *23*(24), 4072–4079. <https://doi.org/10.1002/elps.200290023>
- Ullah, M., Akbar, A., Ng, N. N., Concepcion, W., & Thakor, A. S. (2019). Mesenchymal stem cells confer chemoresistance in breast cancer via a CD9 dependent mechanism. *Oncotarget*, *10*(37), 3435–3450. <https://doi.org/10.18632/oncotarget.26952>
- Urbán, N., & Cheung, T. H. (2021). Stem cell quiescence: the challenging path to activation. *Development (Cambridge, England)*, *148*(3), dev165084. <https://doi.org/10.1242/dev.165084>
- Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T., & Nagao, M. (1997). Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proceedings of the National Academy of Sciences*, *94*(6), 2284–2289. <https://doi.org/10.1073/pnas.94.6.2284>

V

- Vaidya, J. S., Tobias, J. S., Baum, M., Keshtgar, M., Joseph, D., Wenz, F., Houghton, J., Saunders, C., Corica, T., D'Souza, D., Sainsbury, R., Massarut, S., Taylor, I., & Hilaris, B. (2004). Intraoperative radiotherapy for breast cancer. *The Lancet Oncology*, *5*(3), 165–173. [https://doi.org/10.1016/s1470-2045\(04\)01412-3](https://doi.org/10.1016/s1470-2045(04)01412-3)
- Valencia, C. A., Pervaiz, M. A., Husami, A., Qian, Y., & Zhang, K. (2013). Next Generation Sequencing Technologies in Medical Genetics. *SpringerBriefs in Genetics*, 3–11. https://doi.org/10.1007/978-1-4614-9032-6_1
- Vargo-Gogola, T., & Rosen, J. M. (2007). Modelling breast cancer: one size does not fit all. *Nature Reviews Cancer*, *7*(9), 659–672. <https://doi.org/10.1038/nrc2193>
- Varley, J. M., Swallow, J. E., Brammar, W. J., Whittaker, J. L., & Walker, R. A. (1987). Alterations to either c-erbB-2(neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. *Oncogene*, *1*(4), 423–30.
- Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature*, *575*(7782), 299–309. <https://doi.org/10.1038/s41586-019-1730-1>
- Vavouri, T., & Lehner, B. (2012). Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biology*, *13*(11), R110. <https://doi.org/10.1186/gb-2012-13-11-r110>
- Vaziri, N., Shariati, L., Zarrabi, A., Farazmand, A., & Javanmard, S. H. (2021). Cancer-Associated Fibroblasts Regulate the Plasticity of Breast Cancer Stemness through the Production of Leukemia Inhibitory Factor. *Life*, *11*(12), 1298. <https://doi.org/10.3390/life11121298>
- Veer, L. J. V., & Weigelt, B. (2003). Road map to metastasis. *Nature Medicine*, *9*(8), 999–1000. <https://doi.org/10.1038/nm0803-999b>
- Veillard, A.-C., Datlinger, P., Laczik, M., Squazzo, S., & Bock, C. (2016). Diagenode® Premium RRBS technology: cost-effective DNA methylation mapping with superior coverage. *Nature Methods*, *13*(2), i–ii. <https://doi.org/10.1038/nmeth.f.391>
- Velásquez, M. M., Gómez-Maquet, Y., Ferro, E., Cárdenas, W., González-Nieves, S., & Lattig, M. C. (2021). Multidimensional Analysis of Major Depression: Association Between BDNF Methylation, Psychosocial and Cognitive Domains. *Frontiers in Psychiatry*, *12*, 768680. <https://doi.org/10.3389/fpsy.2021.768680>
- Viale, A., Franco, F. D., Orleth, A., Cambiaghi, V., Giuliani, V., Bossi, D., Ronchini, C., Ronzoni, S., Muradore, I., Monestiroli, S., Gobbi, A., Alcalay, M., Minucci, S., & Pelicci, P. G. (2009). Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. *Nature*, *457*(7225), 51–56. <https://doi.org/10.1038/nature07618>
- Vijver, M. J. v. d., Peterse, J. L., Mooi, W. J., Wisman, P., Lomans, J., Dalesio, O., & Nusse, R. (1988). Neu-Protein Overexpression in Breast Cancer. *The New England Journal of Medicine*, *319*(19), 1239–1245. <https://doi.org/10.1056/nejm198811103191902>
- Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Eynde, A. V., Bernard, D., Vanderwinden, J.-M., Bollen, M., Esteller, M., Croce, L. D., Launoit, Y. d., & Fuks, F. (2005). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*, *439*(7078), 871–874. <https://doi.org/10.1038/nature04431>
- Vispé, S., Deroide, A., Davoine, E., Desjobert, C., Lestienne, F., Fournier, L., Novosad, N., Bréand, S., Besse, J., Busato, F., Tost, J., Vries, L. D., Cussac, D., Rioud, J., & Arimondo, P. B. (2015). Consequences of combining siRNA-mediated DNA methyltransferase 1 depletion with 5-aza-2'-deoxycytidine in human leukemic KG1 cells. *Oncotarget*, *6*(17), 15265–15282. <https://doi.org/10.18632/oncotarget.3317>
- Visvader, J. E. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes & Development*, *23*(22), 2563–2577. <https://doi.org/10.1101/gad.1849509>
- Visvader, J. E. (2011). Cells of origin in cancer. *Nature*, *469*(7330), 314–322. <https://doi.org/10.1038/nature09781>
- Visvader, J. E., & Lindeman, G. J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature Reviews Cancer*, *8*(10), 755–768. <https://doi.org/10.1038/nrc2499>
- Vlashi, E., Lagadec, C., Chan, M., Frohnen, P., McDonald, A. J., & Pajonk, F. (2013). Targeted elimination of breast cancer cells with low proteasome activity is sufficient for tumor regression. *Breast Cancer Research and Treatment*, *141*(2), 197–203. <https://doi.org/10.1007/s10549-013-2688-6>
- Vlashi, E., Lagadec, C., Vergnes, L., Reue, K., Frohnen, P., Chan, M., Alhiyari, Y., Dratver, M. B., & Pajonk, F. (2014). Metabolic differences in breast cancer stem cells and differentiated progeny. *Breast Cancer Research and Treatment*, *146*(3), 525–534. <https://doi.org/10.1007/s10549-014-3051-2>
- Vlashi, E., McBride, W. H., & Pajonk, F. (2009). Radiation responses of cancer stem cells. *Journal of Cellular Biochemistry*, *108*(2), 339–342. <https://doi.org/10.1002/jcb.22275>
- Vlerken, L. E. v., Kiefer, C. M., Morehouse, C., Li, Y., Groves, C., Wilson, S. D., Yao, Y., Hollingsworth, R. E., & Hurt, E. M. (2013). EZH2 Is Required for Breast and Pancreatic Cancer Stem Cell Maintenance and Can Be Used as a Functional Cancer Stem Cell Reporter. *STEM CELLS Translational Medicine*, *2*(1), 43–52. <https://doi.org/10.5966/sctm.2012-0036>

Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L., & Frank-Kamenetskii, M. D. (1984). Allowance for Heterogeneous Stacking in the DNA Helix-Coil Transition Theory. *Journal of Biomolecular Structure and Dynamics*, 2(1), 131–148. <https://doi.org/10.1080/07391102.1984.10507552>

Voss, K. O., Roos, K. P., Nonay, R. L., & Dovichi, N. J. (1998). Combating PCR Bias in Bisulfite-Based Cytosine Methylation Analysis. Betaine-Modified Cytosine Deamination PCR. *Analytical Chemistry*, 70(18), 3818–3823. <https://doi.org/10.1021/ac980067t>

W

Wainwright, E. N., & Scaffidi, P. (2017). Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. *Trends in Cancer*, 3(5), 372–386. <https://doi.org/10.1016/j.trecan.2017.04.004>

Walcher, L., Kistenmacher, A.-K., Suo, H., Kitte, R., Dluczek, S., Strauß, A., Blaudszun, A.-R., Yevsa, T., Fricke, S., & Kossatz-Boehlert, U. (2020). Cancer Stem Cells—Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. *Frontiers in Immunology*, 11, 1280. <https://doi.org/10.3389/fimmu.2020.01280>

Wang, D.-Y., Jiang, Z., Ben-David, Y., Woodgett, J. R., & Zacksenhaus, E. (2019a). Molecular stratification within triple-negative breast cancer subtypes. *Scientific Reports*, 9(1), 19107. <https://doi.org/10.1038/s41598-019-55710-w>

Wang, H., Yang, Y., Liu, J., & Qian, L. (2021). Direct cell reprogramming: approaches, mechanisms and progress. *Nature Reviews Molecular Cell Biology*, 22(6), 410–424. <https://doi.org/10.1038/s41580-021-00335-z>

Wang, H., Wang, X., Xu, X., Kyba, M., & Cooney, A. J. (2016). Germ Cell Nuclear Factor (GCNF) Represses Oct4 Expression and Globally Modulates Gene Expression in Human Embryonic Stem (hES) Cells*. *The Journal of Biological Chemistry*, 291(16), 8644–8652. <https://doi.org/10.1074/jbc.m115.694208>

Wang, H., Yu, K., Hou, J., Liu, Q., & Han, W. (2015a). Radiation-induced bystander effect: Early process and rapid assessment. *Cancer Letters*, 356(1), 137–144. <https://doi.org/10.1016/j.canlet.2013.09.031>

Wang, H., Wang, L., Song, Y., Wang, S., Huang, X., Xuan, Q., Kang, X., & Zhang, Q. (2017a). CD44+/CD24- phenotype predicts a poor prognosis in triple-negative breast cancer. *Oncology Letters*, 14(5), 5890–5898. <https://doi.org/10.3892/ol.2017.6959>

Wang, J., Yang, X., Ruan, B., Dai, B., Gao, Y., Duan, J., Qu, S., Tao, K., Dou, K., & Li, H. (2015b). Overexpression of miR-200a suppresses epithelial-mesenchymal transition of liver cancer stem cells. *Tumor Biology*, 36(4), 2447–2456. <https://doi.org/10.1007/s13277-014-2856-2>

Wang, J., Hevi, S., Kurash, J. K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E., & Chen, T. (2009a). The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genetics*, 41(1), 125–129. <https://doi.org/10.1038/ng.268>

Wang, K., Chen, Y., Chang, E.-A., Knott, J. G., & Cibelli, J. B. (2009b). Dynamic Epigenetic Regulation of the Oct4 and Nanog Regulatory Regions during Neural Differentiation in Rhesus Nuclear Transfer Embryonic Stem Cells. *Cloning and Stem Cells*, 11(4), 483–496. <https://doi.org/10.1089/clo.2009.0019>

Wang, L., Huang, X., Zheng, X., Wang, X., Li, S., Zhang, L., Yang, Z., & Xia, Z. (2013a). Enrichment of Prostate Cancer Stem-Like Cells from Human Prostate Cancer Cell Lines by Culture in Serum-Free Medium and Chemoradiotherapy. *International Journal of Biological Sciences*, 9(5), 472–479. <https://doi.org/10.7150/ijbs.5855>

Wang, L., Jia, Y., Jiang, Z., Gao, W., & Wang, B. (2017b). FSCN1 is upregulated by SNAI2 and promotes epithelial to mesenchymal transition in head and neck squamous cell carcinoma. *Cell Biology International*, 41(8), 833–841. <https://doi.org/10.1002/cbin.10786>

Wang, P., Gong, S., Liao, B., Pan, J., Wang, J., Zou, D., Zhao, L., Xiong, S., Deng, Y., Yan, Q., & Wu, N. (2022). HIF1 α /HIF2 α induces glioma cell dedifferentiation into cancer stem cells through Sox2 under hypoxic conditions. *Journal of Cancer*, 13(1), 1–14. <https://doi.org/10.7150/jca.54402>

Wang, P., Lan, C., Xiong, S., Zhao, X., Shan, Y., Hu, R., Wan, W., Yu, S., Liao, B., Li, G., Wang, J., Zou, D., Chen, B., Feng, H., & Wu, N. (2017c). HIF1 α regulates single differentiated glioma cell dedifferentiation to stem-like cell phenotypes with high tumorigenic potential under hypoxia. *Oncotarget*, 8(17), 28074–28092. <https://doi.org/10.18632/oncotarget.15888>

Wang, P., Wan, W.-w., Xiong, S.-L., Feng, H., & Wu, N. (2017d). Cancer stem-like cells can be induced through dedifferentiation under hypoxic conditions in glioma, hepatoma and lung cancer. *Cell Death Discovery*, 3(1), 16105. <https://doi.org/10.1038/cddiscovery.2016.105>

Wang, R. Y.-H., Gehrke, C. W., & Ehrlich, M. (1980). Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Research*, 8(20), 4777–4790. <https://doi.org/10.1093/nar/8.20.4777>

Wang, T., Liu, Q., Li, X., Wang, X., Li, J., Zhu, X., Sun, Z. S., & Wu, J. (2013b). RRBS-Analyser: A Comprehensive Web Server for Reduced Representation Bisulfite Sequencing Data Analysis. *Human Mutation*, 34(12), 1606–1610. <https://doi.org/10.1002/humu.22444>

- Wang, X. Q., Ng, R. K., Ming, X., Zhang, W., Chen, L., Chu, A. C. Y., Pang, R., Lo, C. M., Tsao, S. W., Liu, X., Poon, R. T. P., & Fan, S. T. (2013c). Epigenetic Regulation of Pluripotent Genes Mediates Stem Cell Features in Human Hepatocellular Carcinoma and Cancer Cell Lines. *PLoS ONE*, *8*(9), e72435. <https://doi.org/10.1371/journal.pone.0072435>
- Wang, X., Zhang, H., & Chen, X. (2019b). Drug resistance and combating drug resistance in cancer. *Cancer Drug Resistance*, *2*(2), 141–160. <https://doi.org/10.20517/cdr.2019.10>
- Wang, Y., Li, W., Patel, S. S., Cong, J., Zhang, N., Sabbatino, F., Liu, X., Qi, Y., Huang, P., Lee, H., Taghian, A., Li, J.-J., DeLeo, A. B., Ferrone, S., Epperly, M. W., Ferrone, C. R., Ly, A., Brachtel, E. F., & Wang, X. (2014a). Blocking the formation of radiation-induced breast cancer stem cells. *Oncotarget*, *5*(11), 3743–3755. <https://doi.org/10.18632/oncotarget.1992>
- Wang, Y., Cardenas, H., Fang, F., Condello, S., Taverna, P., Segar, M., Liu, Y., Nephew, K. P., & Matei, D. (2014b). Epigenetic Targeting of Ovarian Cancer Stem Cells. *Cancer Research*, *74*(17), 4922–4936. <https://doi.org/10.1158/0008-5472.can-14-1022>
- Wang, Y., Feng, Q., Ji, C., Liu, X., Li, L., & Luo, J. (2017e). RUNX3 plays an important role in mediating the BMP9-induced osteogenic differentiation of mesenchymal stem cells. *International Journal of Molecular Medicine*, *40*(6), 1991–1999. <https://doi.org/10.3892/ijmm.2017.3155>
- Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., & Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Research*, *25*(21), 4422–4426. <https://doi.org/10.1093/nar/25.21.4422>
- Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., & Clark, S. J. (2002). Identification and resolution of artifacts in bisulfite sequencing. *Methods*, *27*(2), 101–107. [https://doi.org/10.1016/s1046-2023\(02\)00060-9](https://doi.org/10.1016/s1046-2023(02)00060-9)
- Wawruszak, A., Kalafut, J., Okon, E., Czapinski, J., Halasa, M., Przybyszewska, A., Miziak, P., Okla, K., Rivero-Muller, A., & Stepulak, A. (2019). Histone Deacetylase Inhibitors and Phenotypical Transformation of Cancer Cells. *Cancers*, *11*(2), 148. <https://doi.org/10.3390/cancers11020148>
- Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., & Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, *37*(8), 853–862. <https://doi.org/10.1038/ng1598>
- Weigelt, B., Baehner, F. L., & Reis-Filho, J. S. (2010a). The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *The Journal of Pathology*, *220*(2), 263–280. <https://doi.org/10.1002/path.2648>
- Weigelt, B., Mackay, A., A'hern, R., Natrajan, R., Tan, D. S., Dowsett, M., Ashworth, A., & Reis-Filho, J. S. (2010b). Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *The Lancet Oncology*, *11*(4), 339–349. [https://doi.org/10.1016/s1470-2045\(10\)70008-5](https://doi.org/10.1016/s1470-2045(10)70008-5)
- Weigelt, B., Peterse, J. L., & Veer, L. J. v. (2005). Breast cancer metastasis: markers and models. *Nature Reviews Cancer*, *5*(8), 591–602. <https://doi.org/10.1038/nrc1670>
- Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., Hausen, A. z., Brunton, V. G., Morton, J., Sansom, O., Schüler, J., Stemmler, M. P., Herzberger, C., Hopt, U., Keck, T., Brabletz, S., & Brabletz, T. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nature Cell Biology*, *11*(12), 1487–1495. <https://doi.org/10.1038/ncb1998>
- Wen, Y., Cai, J., Hou, Y., Huang, Z., & Wang, Z. (2015). Role of EZH2 in cancer stem cells: from biological insight to a therapeutic target. *Oncotarget*, *8*(23), 37974–37990. <https://doi.org/10.18632/oncotarget.16467>
- Wen, Y., Hou, Y., Yi, X., Sun, S., Guo, J., He, X., Li, T., Cai, J., & Wang, Z. (2021). EZH2 activates CHK1 signaling to promote ovarian cancer chemoresistance by maintaining the properties of cancer stem cells. *Theranostics*, *11*(4), 1795–1813. <https://doi.org/10.7150/thno.48101>
- Wesbuer, S., Lanvers-Kaminsky, C., Duran-Seuberth, I., Bölling, T., Schäfer, K.-L., Braun, Y., Willich, N., & Greve, B. (2010). Association of telomerase activity with radio- and chemosensitivity of neuroblastomas. *Radiation Oncology (London, England)*, *5*(1), 66–66. <https://doi.org/10.1186/1748-717x-5-66>
- Wijaya, A. Y. (2019). Cellular Plasticity and Dedifferentiation: A Link Between Cancer Stem Cells, Hypoxia, Cell Injury, and Inflammation. *Journal of Stem Cell Research and Tissue Engineering*, *2*(2). <https://doi.org/10.20473/jscrte.v2i2.11655>
- Winiiecki, J. (2020). Principles of radiation therapy. *Physical Sciences Reviews*, *0*(0), 20190063. <https://doi.org/10.1515/psr-2019-0063>
- Winslow, T. (n.d.). Medical And Scientific Illustration, Breast, Terese Winslow LLC. <https://www.teresewinslow.com/%5C#/breast/>

- Withers, H. R. (1975). The Four R's of Radiotherapy. *Advances in Radiation Biology*, 5, 241–271. <https://doi.org/10.1016/b978-0-12-035405-4.50012-8>
- Wojdacz, T. K., Borgbo, T., & Hansen, L. L. (2009). Primer design versus PCR bias in methylation independent PCR amplifications. *Epigenetics*, 4(4), 231–234. <https://doi.org/10.4161/epi.9020>
- Wojdacz, T. K., & Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Research*, 35(6), e41–e41. <https://doi.org/10.1093/nar/gkm013>
- Wojdacz, T. K., & Hansen, L. L. (2006). Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. *BioTechniques*, 41(3), 274–278. <https://doi.org/10.2144/000112240>
- Wongtrakoongate, P. (2015). Epigenetic therapy of cancer stem and progenitor cells by targeting DNA methylation machineries. *World Journal of Stem Cells*, 7(1), 137. <https://doi.org/10.4252/wjsc.v7.i1.137>
- Worm, J., Aggerholm, A., & Guldborg, P. (2001). In-Tube DNA Methylation Profiling by Fluorescence Melting Curve Analysis. *Clinical Chemistry*, 47(7), 1183–1189. <https://doi.org/10.1093/clinchem/47.7.1183>
- Wreczycka, K., Gosdschan, A., Yusuf, D., Grüning, B., Assenov, Y., & Akalin, A. (2017). Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology*, 261, 105–115. <https://doi.org/10.1016/j.jbiotec.2017.08.007>
- Wu, C., Guo, E., Ming, J., Sun, W., Nie, X., Sun, L., Peng, S., Luo, M., Liu, D., Zhang, L., Mei, Q., Long, G., Hu, G., & Hu, G. (2020). Radiation-Induced DNMT3B Promotes Radioresistance in Nasopharyngeal Carcinoma through Methylation of p53 and p21. *Molecular Therapy Oncolytics*, 17, 306–319. <https://doi.org/10.1016/j.omto.2020.04.007>
- Wu, C., & Alman, B. A. (2008). Side population cells in human cancers. *Cancer Letters*, 268(1), 1–9. <https://doi.org/10.1016/j.canlet.2008.03.048>
- Wuerstein, R., & Harbeck, N. (2017). Neoadjuvant Therapy for HER2-positive Breast Cancer. *Reviews on Recent Clinical Trials*, 12(2), 81–92. <https://doi.org/10.2174/1574887112666170202165049>

X

- Xavier, M. J., Roman, S. D., Aitken, R. J., & Nixon, B. (2019). Transgenerational inheritance: how impacts to the epigenetic and genetic information of parents affect offspring health. *Human Reproduction Update*, 25(5), 519–541. <https://doi.org/10.1093/humupd/dmz017>
- Xiao, W., Zheng, S., Xie, X., Li, X., Zhang, L., Yang, A., Wang, J., Tang, H., & Xie, X. (2020). SOX2 Promotes Brain Metastasis of Breast Cancer by Upregulating the Expression of FSCN1 and HBEGF. *Molecular Therapy - Oncolytics*, 17, 118–129. <https://doi.org/10.1016/j.omto.2020.03.001>
- Xie, G., Yao, Q., Liu, Y., Du, S., Liu, A., Guo, Z., Sun, A., Ruan, J., Chen, L., Ye, C., & Yuan, Y. (2011). IL-6-induced epithelial-mesenchymal transition promotes the generation of breast cancer stem-like cells analogous to mammosphere cultures. *International Journal of Oncology*, 40(4), 1171–1179. <https://doi.org/10.3892/ijo.2011.1275>
- Xie, J., Xiao, Y., Zhu, X.-y., Ning, Z.-y., Xu, H.-f., & Wu, H.-m. (2016). Hypoxia regulates stemness of breast cancer MDA-MB-231 cells. *Medical Oncology (Northwood, London, England)*, 33(5), 42. <https://doi.org/10.1007/s12032-016-0755-7>
- Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, 1(3), e90. <https://doi.org/10.1002/cpz1.90>
- Xiong, Z., & Laird, P. W. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Research*, 25(12), 2532–2534. <https://doi.org/10.1093/nar/25.12.2532>
- Xu, B., Shen, J., Guo, W., Zhao, W., Zhuang, Y., & Wang, L. (2019). Impact of the 2018 ASCO/CAP HER2 guidelines update for HER2 testing by FISH in breast cancer. *Pathology - Research and Practice*, 215(2), 251–255. <https://doi.org/10.1016/j.prp.2018.10.035>
- Xu, J., Liao, K., & Zhou, W. (2018). Exosomes Regulate the Transformation of Cancer Cells in Cancer Stem Cell Homeostasis. *Stem Cells International*, 2018, 4837370. <https://doi.org/10.1155/2018/4837370>
- Xu, J., Chen, Y., & Olopade, O. I. (2010). MYC and Breast Cancer. *Genes & Cancer*, 1(6), 629–640. <https://doi.org/10.1177/1947601910378691>
- Xu, S., Wang, J., Ding, N., Hu, W., Zhang, X., Wang, B., Hua, J., Wei, W., & Zhu, Q. (2016). Exosome-mediated microRNA transfer plays a role in radiation-induced bystander effect. *RNA Biology*, 12(12), 1355–1363. <https://doi.org/10.1080/15476286.2015.1100795>
- Xu, X., Chai, S., Wang, P., Zhang, C., Yang, Y., Yang, Y., & Wang, K. (2015). Aldehyde dehydrogenases and cancer stem cells. *Cancer Letters*, 369(1), 50–57. <https://doi.org/10.1016/j.canlet.2015.08.018>

Y

- Yadav, P., & Shankar, B. S. (2019). Radio resistance in breast cancer cells is mediated through TGF- β signalling, hybrid epithelial-mesenchymal phenotype and cancer stem cells. *Biomedicine & Pharmacotherapy*, *111*(Br. J. Radiol. 86 1023 2013), 119–130. <https://doi.org/10.1016/j.biopha.2018.12.055>
- Yakovchuk, P., Protozanova, E., & Frank-Kamenetskii, M. D. (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*, *34*(2), 564–574. <https://doi.org/10.1093/nar/gkj454>
- Yamada, Y., Haga, H., & Yamada, Y. (2014). Concise Review: Dedifferentiation Meets Cancer Development: Proof of Concept for Epigenetic Cancer. *STEM CELLS Translational Medicine*, *3*(10), 1182–1187. <https://doi.org/10.5966/sctm.2014-0090>
- Yamashita, N., Tokunaga, E., Kitao, H., Hitchins, M., Inoue, Y., Tanaka, K., Hisamatsu, Y., Taketani, K., Akiyoshi, S., Okada, S., Oda, Y., Saeki, H., Oki, E., & Maehara, Y. (2015). Epigenetic Inactivation of BRCA1 Through Promoter Hypermethylation and Its Clinical Importance in Triple-Negative Breast Cancer. *Clinical Breast Cancer*, *15*(6), 498–504. <https://doi.org/10.1016/j.clbc.2015.06.009>
- Yamashita, Y. M., & Fuller, M. T. (2008). Asymmetric centrosome behavior and the mechanisms of stem cell division. *The Journal of Cell Biology*, *180*(2), 261–266. <https://doi.org/10.1083/jcb.200707083>
- Yamazaki, H., Xu, C. W., Naito, M., Nishida, H., Okamoto, T., Ghani, F. I., Iwata, S., Inukai, T., Sugita, K., & Morimoto, C. (2011). Regulation of cancer stem cell properties by CD9 in human B-acute lymphoblastic leukemia. *Biochemical and Biophysical Research Communications*, *409*(1), 14–21. <https://doi.org/10.1016/j.bbrc.2011.04.098>
- Yamazaki, J., Estecio, M. R., Lu, Y., Long, H., Malouf, G. G., Graber, D., Huo, Y., Ramagli, L., Liang, S., Kornblau, S. M., Jelinek, J., & Issa, J.-P. J. (2012). The epigenome of AML stem and progenitor cells. *Epigenetics*, *8*(1), 92–104. <https://doi.org/10.4161/epi.23243>
- Yang, A. S., Estécio, M. R. H., Doshi, K., Kondo, Y., Tajara, E. H., & Issa, J.-P. J. (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Research*, *32*(3), e38–e38. <https://doi.org/10.1093/nar/gnh032>
- Yang, G., Quan, Y., Wang, W., Fu, Q., Wu, J., Mei, T., Li, J., Tang, Y., Luo, C., Ouyang, Q., Chen, S., Wu, L., Hei, T. K., & Wang, Y. (2012). Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. *British Journal of Cancer*, *106*(9), 1512–1519. <https://doi.org/10.1038/bjc.2012.126>
- Yang, J., Liao, D., Chen, C., Liu, Y., Chuang, T.-H., Xiang, R., Markowitz, D., Reisfeld, R. A., & Luo, Y. (2013). Tumor-Associated Macrophages Regulate Murine Breast Cancer Stem Cells Through a Novel Paracrine EGFR/Stat3/Sox-2 Signaling Pathway. *STEM CELLS*, *31*(2), 248–258. <https://doi.org/10.1002/stem.1281>
- Yang, L., Shi, P., Zhao, G., Xu, J., Peng, W., Zhang, J., Zhang, G., Wang, X., Dong, Z., Chen, F., & Cui, H. (2020a). Targeting cancer stem cell pathways for cancer therapy. *Signal Transduction and Targeted Therapy*, *5*(1), 8. <https://doi.org/10.1038/s41392-020-0110-5>
- Yang, L., Tang, H., Kong, Y., Xie, X., Chen, J., Song, C., Liu, X., Ye, F., Li, N., Wang, N., & Xie, X. (2015). LGR5 Promotes Breast Cancer Progression and Maintains Stem-Like Cells Through Activation of Wnt/ β -Catenin Signaling. *STEM CELLS*, *33*(10), 2913–2924. <https://doi.org/10.1002/stem.2083>
- Yang, Q., Zhao, S., Shi, Z., Cao, L., Liu, J., Pan, T., Zhou, D., & Zhang, J. (2021). Chemotherapy-elicited exosomal miR-378a-3p and miR-378d promote breast cancer stemness and chemoresistance via the activation of EZH2/STAT3 signaling. *Journal of Experimental & Clinical Cancer Research : CR*, *40*(1), 120. <https://doi.org/10.1186/s13046-021-01901-1>
- Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., & Liang, G. (2014). Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell*, *26*(4), 577–590. <https://doi.org/10.1016/j.ccr.2014.07.028>
- Yang, X., Lin, X., Zhong, X., Kaur, S., Li, N., Liang, S., Lassus, H., Wang, L., Katsaros, D., Montone, K., Zhao, X., Zhang, Y., Bützow, R., Coukos, G., & Zhang, L. (2010). Double-Negative Feedback Loop between Reprogramming Factor LIN28 and microRNA let-7 Regulates Aldehyde Dehydrogenase 1-Positive Cancer Stem Cells. *Cancer Research*, *70*(22), 9463–9472. <https://doi.org/10.1158/0008-5472.can-10-2388>
- Yang, Y., Li, X., Wang, T., Guo, Q., Xi, T., & Zheng, L. (2020b). Emerging agents that target signaling pathways in cancer stem cells. *Journal of Hematology & Oncology*, *13*(1), 60. <https://doi.org/10.1186/s13045-020-00901-6>
- Yarden, Y., & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology*, *2*(2), 127–137. <https://doi.org/10.1038/35052073>
- Yeo, S., Jeong, S., Kim, J., Han, J.-S., Han, Y.-M., & Kang, Y.-K. (2007). Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells. *Biochemical and Biophysical Research Communications*, *359*(3), 536–542. <https://doi.org/10.1016/j.bbrc.2007.05.120>
- Yersal, O., & Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World Journal of Clinical Oncology*, *5*(3), 412. <https://doi.org/10.5306/wjco.v5.i3.412>

- Yi, J. M., Tsai, H.-C., Glöckner, S. C., Lin, S., Ohm, J. E., Easwaran, H., James, C. D., Costello, J. F., Riggins, G., Eberhart, C. G., Latterra, J., Vescovi, A. L., Ahuja, N., Herman, J. G., Schuebel, K. E., & Baylin, S. B. (2008). Abnormal DNA Methylation of CD133 in Colorectal and Glioblastoma Tumors. *Cancer Research*, *68*(19), 8094–8103. <https://doi.org/10.1158/0008-5472.can-07-6208>
- Yilmaz, M., & Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. *Cancer and Metastasis Reviews*, *28*(1-2), 15–33. <https://doi.org/10.1007/s10555-008-9169-0>
- Yin, X., Zhang, B.-H., Zheng, S.-S., Gao, D.-M., Qiu, S.-J., Wu, W.-Z., & Ren, Z.-G. (2015). Coexpression of gene Oct4 and Nanog initiates stem cell characteristics in hepatocellular carcinoma and promotes epithelial-mesenchymal transition through activation of Stat3/Snail signaling. *Journal of Hematology & Oncology*, *8*(1), 23. <https://doi.org/10.1186/s13045-015-0119-3>
- You, H., Ding, W., & Rountree, C. B. (2010). Epigenetic regulation of cancer stem cell marker CD133 by transforming growth factor- β . *Hepatology*, *51*(5), 1635–1644. <https://doi.org/10.1002/hep.23544>
- Young, J., Wu, S., Hansteen, G., Du, C., Sambucetti, L., Remiszewski, S., O'Farrell, A.-M., Hill, B., Lavau, C., & Murray, L. (2004). Inhibitors of histone deacetylases promote hematopoietic stem cell self-renewal. *Cytotherapy*, *6*(4), 328–336. <https://doi.org/10.1080/14653240410004899>
- Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., & Song, E. (2007a). let-7 Regulates Self Renewal and Tumorigenicity of Breast Cancer Cells. *Cell*, *131*(6), 1109–1123. <https://doi.org/10.1016/j.cell.2007.10.054>
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2007b). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*, *318*(5858), 1917–1920. <https://doi.org/10.1126/science.1151526>
- Yu, X., & Sun, S. (2016). HMM-DM: identifying differentially methylated regions using a hidden Markov model. *Statistical Applications in Genetics and Molecular Biology*, *15*(1), 69–81. <https://doi.org/10.1515/sagmb-2015-0077>
- Yu, X., Teng, Y., Jiang, X., Yuan, H., & Jiang, W. (2020). Genome-Wide DNA Methylation Pattern of Cancer Stem Cells in Esophageal Cancer. *Technology in Cancer Research & Treatment*, *19*, 1533033820983793. <https://doi.org/10.1177/1533033820983793>
- Yuan, S., Norgard, R. J., & Stanger, B. Z. (2019). Cellular Plasticity in Cancer. *Cancer Discovery*, *9*(7), 837–851. <https://doi.org/10.1158/2159-8290.cd-19-0015>
- Yuanxiang, Z., Magill, J. M., Newton, R., & Magill, C. (1997). Use of a Single Sequencing Termination Reaction to Distinguish Between Cytosine and 5-Methylcytosine in Bisulfite-Modified DNA. *BioTechniques*, *22*(5), 850–854. <https://doi.org/10.2144/97225bm14>
- ## Z
- Zagorac, S., Alcala, S., Bayon, G. F., Kheir, T. B., Schoenhals, M., González-Neira, A., Fraga, M. F., Aicher, A., Heeschen, C., & Sainz, B. (2016). DNMT1 Inhibition Reprograms Pancreatic Cancer Stem Cells via Upregulation of the miR-17-92 Cluster. *Cancer Research*, *76*(15), 4546–4558. <https://doi.org/10.1158/0008-5472.can-15-3268>
- Zang, Q., Xu, L., Li, J., & Jia, H. (2020). GATA6 activated long non-coding RNA PCAT1 maintains stemness of non-small cell lung cancer by mediating FRK. *Journal of B.U.ON. : official journal of the Balkan Union of Oncology*, *25*(5), 2371–2381.
- Zanoni, M., Bravaccini, S., Fabbri, F., & Arienti, C. (2022). Emerging Roles of Aldehyde Dehydrogenase Isoforms in Anti-cancer Therapy Resistance. *Frontiers in Medicine*, *9*, 795762. <https://doi.org/10.3389/fmed.2022.795762>
- Zapperi, S., & Porta, C. A. M. L. (2012). Do cancer cells undergo phenotypic switching? The case for imperfect cancer stem cell markers. *Scientific Reports*, *2*(1), 441. <https://doi.org/10.1038/srep00441>
- Zeng, Y., & Chen, T. (2019). DNA Methylation Reprogramming during Mammalian Development. *Genes*, *10*(4), 257. <https://doi.org/10.3390/genes10040257>
- Zhan, H.-x., Wang, Y., Li, C., Xu, J.-w., Zhou, B., Zhu, J.-k., Han, H.-f., Wang, L., Wang, Y.-s., & Hu, S.-y. (2016). LincRNA-ROR promotes invasion, metastasis and tumor growth in pancreatic cancer through activating ZEB1 pathway. *Cancer Letters*, *374*(2), 261–271. <https://doi.org/10.1016/j.canlet.2016.02.018>
- Zhang, H., Wu, H., Zheng, J., Yu, P., Xu, L., Jiang, P., Gao, J., Wang, H., & Zhang, Y. (2013a). Transforming Growth Factor β 1 Signal is Crucial for Dedifferentiation of Cancer Cells to Cancer Stem Cells in Osteosarcoma. *STEM CELLS*, *31*(3), 433–446. <https://doi.org/10.1002/stem.1298>
- Zhang, H., Shang, Y.-P., Chen, H.-y., & Li, J. (2017a). Histone deacetylases function as novel potential therapeutic targets for cancer. *Hepatology Research*, *47*(2), 149–159. <https://doi.org/10.1111/hepr.12757>
- Zhang, L., Ye, S.-B., Ma, G., Tang, X.-F., Chen, S.-P., He, J., Liu, W.-L., Xie, D., Zeng, Y.-X., & Li, J. (2013b). The expressions of MIF and CXCR4 protein in tumor microenvironment are adverse prognostic factors in patients with esophageal

- squamous cell carcinoma. *Journal of Translational Medicine*, 11(1), 60–60. <https://doi.org/10.1186/1479-5876-11-60>
- Zhang, L., Shi, H., Chen, H., Gong, A., Liu, Y., Song, L., Xu, X., You, T., Fan, X., Wang, D., Cheng, F., & Zhu, H. (2019). Dedifferentiation process driven by radiotherapy-induced HMGB1/TLR2/YAP/HIF-1 α signaling enhances pancreatic cancer stemness. *Cell Death & Disease*, 10(10), 724. <https://doi.org/10.1038/s41419-019-1956-8>
- Zhang, S., Qin, C., Cao, G., Guo, L., Feng, C., & Zhang, W. (2017b). Genome-wide analysis of DNA methylation profiles in a senescence-accelerated mouse prone 8 brain using whole-genome bisulfite sequencing. *Bioinformatics*, 33(11), btx040. <https://doi.org/10.1093/bioinformatics/btx040>
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E., & Ecker, J. R. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell*, 126(6), 1189–1201. <https://doi.org/10.1016/j.cell.2006.08.003>
- Zhang, Y., Rohde, C., Tierling, S., Stamerjohanns, H., Reinhardt, R., Walter, J., & Jeltsch, A. (2009). DNA Methylation, Methods and Protocols. *Methods in Molecular Biology*, 507, 177–187. https://doi.org/10.1007/978-1-59745-522-0_14
- Zheng, H., Pomyen, Y., Hernandez, M. O., Li, C., Livak, F., Tang, W., Dang, H., Greten, T. F., Davis, J. L., Zhao, Y., Mehta, M., Levin, Y., Shetty, J., Tran, B., Budhu, A., & Wang, X. W. (2018). Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. *Hepatology*, 68(1), 127–140. <https://doi.org/10.1002/hep.29778>
- Zhou, H.-M., Zhang, J.-G., Zhang, X., & Li, Q. (2021). Targeting cancer stem cells for reversing therapy resistance: mechanism, signaling, and prospective agents. *Signal Transduction and Targeted Therapy*, 6(1), 62. <https://doi.org/10.1038/s41392-020-00430-1>
- Zhou, S., Schuetz, J. D., Bunting, K. D., Colapietro, A.-M., Sampath, J., Morris, J. J., Lagutina, I., Grosveld, G. C., Osawa, M., Nakauchi, H., & Sorrentino, B. P. (2001). The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nature Medicine*, 7(9), 1028–1034. <https://doi.org/10.1038/nm0901-1028>
- Zhu, H., Wang, D., Liu, Y., Su, Z., Zhang, L., Chen, F., Zhou, Y., Wu, Y., Yu, M., Zhang, Z., & Shao, G. (2013). Role of the Hypoxia-inducible factor-1 alpha induced autophagy in the conversion of non-stem pancreatic cancer cells into CD133+ pancreatic cancer stem-like cells. *Cancer Cell International*, 13(1), 119–119. <https://doi.org/10.1186/1475-2867-13-119>
- Zhu, X., Wang, Y., Tan, L., & Fu, X. (2018). The pivotal role of DNA methylation in the radio-sensitivity of tumor radiotherapy. *Cancer Medicine*, 7(8), 3812–3819. <https://doi.org/10.1002/cam4.1614>
- Zhu, Z., Chen, X., Xiao, Y., Wen, J., Chen, J., Wang, K., & Chen, G. (2019). Gestational diabetes mellitus alters DNA methylation profiles in pancreas of the offspring mice. *Journal of Diabetes and its Complications*, 33(1), 15–22. <https://doi.org/10.1016/j.jdiacomp.2018.11.002>
- Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., Gnirke, A., & Meissner, A. (2011). Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genetics*, 7(12), e1002389. <https://doi.org/10.1371/journal.pgen.1002389>

APPENDIX 1

PRIMER DESIGN FOR BISULFITE SEQUENCING PCR

≡ CHAPTER CONTENTS

1	PRIMER DESIGN SOFTWARE	305
2	GENERAL FEATURES	306
2.1	DNA TEMPLATE FOR PRIMER DESIGN	306
2.2	PRIMERS LENGTH	307
2.3	AMPLICON LENGTH	308
3	OTHER FEATURES	308
3.1	CPG SITES IN PRIMERS	308
3.2	CONVERTED CYTOSINES IN PRIMERS	309
3.3	PRIMERS GC CONTENT AND T_M DIFFERENCE	310
3.4	PRIMERS BASE COMPOSITION AND COMPLEMENTARITY	310
3.5	PRIMER SPECIFICITY AND UNINTENDED PCR PRODUCTS	310
3.6	AMPLICON BASE COMPOSITION: REPEATS, T_M DIFFERENCE AND CPG CONTENT	311
3.7	ADDITIONAL SEQUENCES	312
4	CONCLUSION	312
	📌 KEY POINTS	313
5	BIBLIOGRAPHY	313

^ Back to Table of Contents

The design of primers is a key step in the experimental conception of BSP assays to ensure a proper estimation of DNA methylation levels. Selecting primers with the right features is essential for the PCR's effectiveness and may prevent or, at least, reduce undesired bias.

1 PRIMER DESIGN SOFTWARE

In the process of designing primers, tools to generate primer pairs, calculate T_m , check for undesired PCR products, manipulate DNA sequences, find genomic coordinates or visualize primer and amplicon regions on the genome can be useful, some of them are listed in [Table A1.1](#). All of these software are only meant to assist researchers in the design of primers to follow, all or most, of the guidelines described below and summarized in [Figure A1.1](#).

Notably, some software are available to facilitate primer generation for methylation-independent PCR (MIP), such as:

- *MethPrimer* (<https://www.urogene.org/methprimer/>, Li, 2007; Li and Dahiya, 2002)

- *BiSearch* (<http://bisearch.enzim.hu/>, Arányi and Tusnády, 2007; Arányi *et al.*, 2006; Tusnády *et al.*, 2005)
- *PRIMEGENS-w3* (<http://primegens.org/>, Kushwaha *et al.*, 2015)
- *PrimerSuite* (<http://www.primer-suite.com/>, Lu *et al.*, 2017)
- *Bisprimer* (<https://www.ibp.cz/local/software/BisPrimer/>, Kovacova and Janousek, 2012)

The *MethBank* (<http://bigd.big.ac.cn/methbank/>) methylation database website from the *National Genomics Data Center* of the *China National Center for Bioinformatics*, provides a collection of software tools, called *MethTools* (<http://bigd.big.ac.cn/methbank/methTool/list/>), listing all the tools related to the analysis of DNA methylation (Li *et al.*, 2018; Zou *et al.*, 2015). All the available software associated to primer design (for MIP/MSP primers for or other DNA methylation assays) can be found by filtering this database by category. There is a total of 13 primer design software referenced for this use.

SOFTWARE	FUNCTIONALITY	WEB LINK	REFERENCES
MethTools	List of software related to DNA methylation study	http://bigd.big.ac.cn/methbank/methTool/list/	Zou <i>et al.</i> , 2015 Li <i>et al.</i> , 2018
MethPrimer, MethPrimer 2.0	Primer design and CpG island prediction	https://www.urogene.org/methprimer/	Li and Dahiya, 2002
BiSearch	Primer design and <i>in-silico</i> PCR for unintended PCR product prediction	http://bisearch.enzim.hu/	Tusnády <i>et al.</i> , 2005 Arányi <i>et al.</i> , 2006 Arányi and Tusnády, 2007
Integrative Genomics Viewer (IGV)	Visualisation on genome, genomic sequence and corresponding coordinates retrieval	https://igv.org/	Thorvaldsdóttir <i>et al.</i> , 2013 Robinson <i>et al.</i> , 2011 Robinson <i>et al.</i> , 2020
BioWord (Microsoft Office Word plugin)	Manipulation of DNA sequences	https://erilllab.umbc.edu/bioword-2/	Anzaldi <i>et al.</i> , 2012
OligoCalc	Oligonucleotide properties calculator, T_m calculation for PCR products	http://biotools.nubic.northwestern.edu/OligoCalc.html	Kibbe, 2007

TABLE A1.1 SOFTWARE THAT AID IN THE PROCESS OF DESIGNING PRIMERS FOR BSP. (Adapted from Hernandez *et al.*, 2013)

2 GENERAL FEATURES

2.1 DNA TEMPLATE FOR PRIMER DESIGN

The bisulfite treatment transforms cytosines in uracils. As guanines of the opposite strand are complementary to cytosines, they cannot bond to uracils, thus creating mismatches: DNA strands are no longer complementary. Thereby, to have functional primer pair, both primers have to be complementary to only one of the original strands, called the template strand in [Figure A1.1](#), and the

regenerated strand complementary to the template strand is referred to as the “virtual” complementary strand (Tusnády *et al.*, 2005). Both of the original strands can be tested as template strands to choose the most optimal primer pair (Ashapkin *et al.*, 2020; Poucke *et al.*, 2017; Tost and Gut, 2007). Indeed, due to the symmetry of CpG motifs and the methyltransferase activity, the methylation status of every CpG site should be identical on both strands, unless the region of interest is known to be found hemimethylated. Thus, the BSP technique is not adapted to study hemimethylation of DNA if only one strand is analyzed.

Due to the substitution of C by T by the bisulfite reaction, the treated DNA presents a high redundancy of Ts, causing several difficulties for PCR, such as a low amplification efficiency, because of long T stretches, as well as frequent primer dimer formation and high frequency of unintended PCR products, because of the bases heterogeneity loss. Therefore, the objective of the primer design when dealing with bisulfite-treated DNA is to adapt primer features to these specificities to ensure an efficient PCR amplification (Arányi and Tusnády, 2007; Parrish *et al.*, 2012).

2.2 PRIMERS LENGTH

In bisulfite-converted DNA, theoretically, all unmethylated cytosines are replaced by thymines. Therefore the DNA sequence is mainly composed of three bases instead of four and the ratio of bases shift from 25% of each base towards 25% of A, 25% of G, 0% of C, and 50% of T, depending on the percentage of unconverted methylated cytosines. Because of this loss of diversity in the base composition, primers designed for bisulfite-treated DNA are not as specific as primers designed for native DNA.

As explained in Poucke *et al.* (2017), the estimation of a primer occurrence in a DNA template can be calculated by the following formula: $N \times pA^{Na} \times pG^{Ng} \times pC^{Nc} \times pT^{Nt}$, with N being the number of nucleotides in the template; pA , pG , pC , and pT being the estimated frequencies of respective nucleotides in the template; and Na , Ng , Nc , and Nt being the numbers of the respective nucleotides in the primer sequence. In a mammalian genome ($\sim 3 \times 10^9$ bp, with 25% of each nucleotide), a native primer of 20 bp (with 25% of each nucleotide, so hypothetically composed of 5 nucleotides of each base) has an occurrence estimated at $3 \times 10^9 \times 0.25^5 \times 0.25^5 \times 0.25^5 \times 0.25^5 = 0.0027$ times, which is considered as highly specific. Whereas, in a bisulfite-treated mammalian genome ($\sim 3 \times 10^9$ bp, considered with 1% of unconverted methylated cytosines (Tost, 2010), so with 25% of A, 25% of G, 1% of C, and 49% of T), a bisulfite primer of 20 bp (with 25% of A, 25% of G, 0% of C, and 50% of T, so hypothetically composed of 5 A, 5 G, and 10 T) has an occurrence estimated at $3 \times 10^9 \times 0.25^5 \times 0.25^5 \times 0.49^{10} = 2.28$ times, which is not enough to be specific only to the targeted sequence.

So, to balance the loss of primer specificity when dealing with bisulfite converted DNA, bisulfite primers for MIP should be longer than native primers, between 25 and 30 bp (Li and Tollefsbol, 2011; Tost and Gut, 2007). Indeed, in a bisulfite-treated mammalian genome, a bisulfite primer of 25 bp

(with 25% of A, 25% of G, 0% of C, and 50% of T, so hypothetically composed of 6 A, 6 G, and 13 T) has an occurrence estimated at $3 \times 10^9 \times 0.25^6 \times 0.25^6 \times 0.49^{13} = 0,017$ times, which is, in theory, considered specific enough. For example, the *MethPrimer* default range for primer size is from 20 to 30 bp with an optimal size being 25 bp, and the *BiSearch* default range is set from 15 to 35 bp (Li and Dahiya, 2002; Tusnády *et al.*, 2005).

2.3 AMPLICON LENGTH

The bisulfite treatment causes strand breakage, therefore it is difficult to amplify fragments longer than 500-600 bp as fewer templates are available (Tanaka and Okamoto, 2007; Warnecke *et al.*, 2002). Thus, PCR product length should not exceed 400 to 500 bp when designing primers for bisulfite-treated DNA. In the literature, it is widely recommended to have a PCR product of 200 to 300 bp. For examples, in Warnecke *et al.* (2002), the authors recommend amplifying a 300 bp fragment, in Tost and Gut (2007) they recommend an optimal PCR product size of 200 bp and no more than 300 bp, and in Chen *et al.* (2017b) the recommended amplicon length is 200 bp and at least less than 500 bp.

3 OTHER FEATURES

3.1 CPG SITES IN PRIMERS

Methylation-independent PCR (MIP) primers should not have any CpG sites in their sequence, allowing them to bind to DNA templates regardless of their CpG methylation status. But, in some cases, it is not possible to find a 20 to 30 bp long sequence without CpG sites in the region of interest (in CpG-rich regions for example) so the presence of a CpG site in a primer sequence is inevitable. Moreover, the incorporation of a CG dinucleotide in primer sequence was also used in numerous studies to reverse the PCR preferential amplification of unmethylated DNA, but it was shown that it can over-correct the bias by enhancing selectivity towards methylated DNA (Candiloro *et al.*, 2017). For that reason, the inclusion of a CpG site in the primer sequence should be approached with great caution and only if no other options are available to amplify the region of interest.

If a CpG site is inevitable in primer sequence, only a maximum of one CpG site should be included and it should be as far as possible from the 3' end or at least not the last 5 nucleotides at the 3' end of primer sequence (Kovacova and Janousek, 2012; Tost and Gut, 2007; Wojdacz and Hansen, 2006; Wojdacz *et al.*, 2008). In the forward primer, a Y must replace the cytosine, meaning it can be either C or T, and in the reverse primer, a R must replace the guanine, meaning it can be either G or A (Candiloro *et al.*, 2017; Chen *et al.*, 2017a; Clark *et al.*, 1994; Poucke *et al.*, 2017; Shen *et al.*, 2007; Warnecke *et al.*, 2002). As, in this case, two sequences of the same primer exists, the fully converted one (for unmethylated CpG) and unconverted one (for methylated CpG), to avoid preferential amplification from one over the other, the T_m difference between these two sequences must be lower than 2.5°C (further details on primer T_m in the section 3.3 "Primers GC content and

T_m difference” at page 310) (Kovacova and Janousek, 2012; Tusnády *et al.*, 2005).

3.2 CONVERTED CYTOSINES IN PRIMERS

As the bisulfite reaction converting cytosines into uracils is never fully efficient, to ensure the specificity of primers towards converted DNA, and not towards unconverted genomic DNA, thymines (or adenines for reverse sequence) resulting from converted cytosines (outside a CpG context) has to be included within the primer sequence. Some publications recommend having at least 4 converted cytosines or that converted cytosines represent a minimum of 25% of all primer bases (Fitzpatrick *et al.*, 1998; Tost and Gut, 2007). Plus, one converted cytosine should be present in the last 5 bases at the 3' end of the primer sequence to further increase the specificity towards converted DNA (Kovacova and Janousek, 2012; Tusnády *et al.*, 2005).

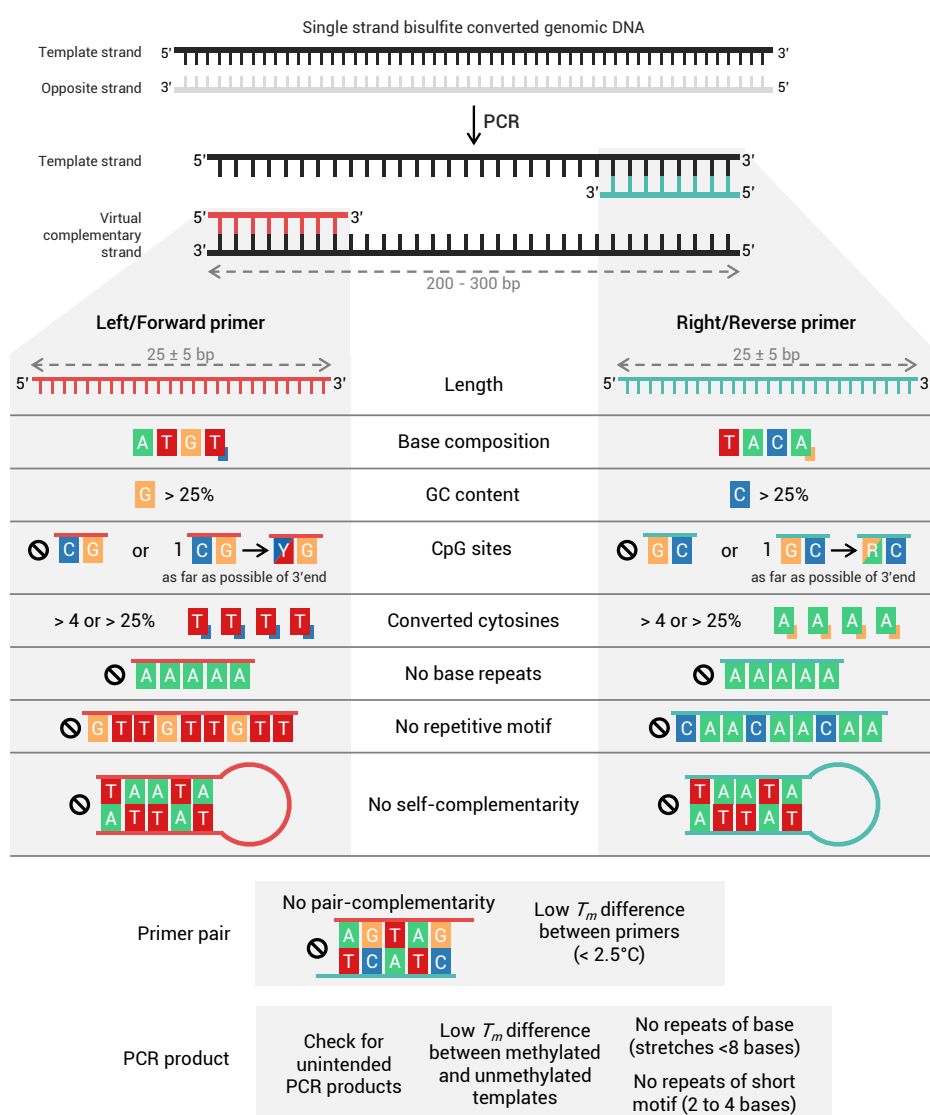


FIGURE A1.1 PRIMER DESIGN GUIDELINES FOR BSP. The different features for primer design are illustrated using examples.

3.3 PRIMERS GC CONTENT AND T_m DIFFERENCE

Because the bisulfite conversion causes the substitution of C by T in the forward primer and G by A in the reverse primer, compared to their original, the G+C ratio over A+T bases is severely reduced, hence it is recommended to choose primers with at least 25% of GC if possible (Fitzpatrick *et al.*, 1998). The default values for optimal GC content set by *MethPrimer* and *BiSearch* are respectively 40% and 30%, and default ranges are respectively 10-80% and 0-60% (Li and Dahiya, 2002; Tusnády *et al.*, 2005). Additionally, the T_m between the forward and reverse primers sequences should be as low as possible, for primers to bind equally on their respective templates at the same annealing temperature during PCR. Indeed, the higher primers T_m differ, the more it increases the risk of non-binding for the lower T_m primer and non-specific binding of the higher T_m primer (Poucke *et al.*, 2017). Hence, having primers with a small T_m difference is a crucial parameter to achieve efficient PCR amplification.

3.4 PRIMERS BASE COMPOSITION AND COMPLEMENTARITY

The primer sequences should not have too long homopolymer strings, in particular long stretches of T or A as it reduces the stability and the specificity of binding (Fitzpatrick *et al.*, 1998; Parrish *et al.*, 2012; Warnecke *et al.*, 2002). However, as thymines have a much higher occurrence in forward primer due to cytosines conversion, repeats of T are quite inevitable but must kept be as short as possible. As an example, the *MethPrimer* software sets a maximum of 5 mononucleotides repeats except for T for which the maximum allowed is 8 (Li and Dahiya, 2002). Moreover, same as for native primers, to have a correct PCR efficacy primers must not form hairpin structures, have repetitive motifs or have high complementarity to each other as they may form primer-dimers (Li and Tollefsbol, 2011; Tost and Gut, 2007; Wojdacz *et al.*, 2008). Thus, the self-complementary score and the 3'-self complementarity score, used to predict possible secondary structures by self-binding or primer dimers formation, have to be as low as possible (Rozen and Skaletsky, 2000; Rychlik, 1995).

3.5 PRIMER SPECIFICITY AND UNINTENDED PCR PRODUCTS

The contamination of PCR products by unintended secondary co-amplified sequences should be avoided as much as possible. Indeed, for cloning-BSP experiments, it leads to incorrect cloned sequences, and for direct-BSP, to superimposed signals resulting in the inability to estimate methylation levels due to mixed nucleotide detection at the CpG positions. Therefore, primers should not bind to and amplify other sequences in the bisulfite-converted DNA template.

As explained in section 2.2 “Primers length” at page 307, the T redundancy in converted DNA causes a loss of primer specificity to the desired region (Arányi and Tusnády, 2007; Parrish *et al.*, 2012). This issue is partially corrected by the increase of primer length to enhance its specificity but it still needs to be verified, to choose the primer pair less susceptible to producing contaminating PCR products.

Among the software listed by *MethTools*, the *BiSearch* software proposes a search tool, called “Primer search, ePCR”, used for post-design prediction of potential PCR products (Arányi and Tusnády, 2007; Arányi *et al.*, 2006; Tusnády *et al.*, 2005). This algorithm performs a similarity search between primers and converted sequences of the selected genome (both strands and virtual complementary strands) (Ashapkin *et al.*, 2020). On the contrary to *Blast* searches (Altschul *et al.*, 1990; Ye *et al.*, 2012), *BiSearch* does not compute the probability of identity of two sequences, but only searches if sequences have the same nucleotides stretches, by comparing base by base the primer sequence (from the 3' end) to the converted genome sequence. Based on the locations and orientations of the hits, the algorithm finds potential PCR products (< 1,000 bp by default) (Arányi and Tusnády, 2007; Arányi *et al.*, 2006; Tusnády *et al.*, 2005).

3.6 AMPLICON BASE COMPOSITION: REPEATS, T_m DIFFERENCE AND CPG CONTENT

When amplifying bisulfite-converted DNA, several artifacts can be resolved by manipulating the amplicon base composition.

Firstly, the presence of long mononucleotide repeats such as poly-A and poly-T or short repetitive motifs can cause polymerase slippage during PCR (Fazekas *et al.*, 2010; Levinson and Gutman, 1987; Sehn, 2015; Shinde *et al.*, 2003). The polymerase slippage generates mutations in PCR products resulting in a misalignment of base signals in the sequencing result, therefore making it difficult to estimate methylation. Hence, stretches of eight or more bases in the PCR product should be avoided, especially for poly-A and poly-T, as well as repetitive short motifs, from one to four bases (Parrish *et al.*, 2012; Sehn, 2015).

Secondly, the PCR amplification of bisulfite-converted DNA can be biased: it can unequally amplify the methylated and unmethylated templates, due to their differences in GC content and melting temperature, as stated in section “PCR bias: unequal amplification of methylated and unmethylated DNA” at page 142 (Guldborg *et al.*, 2002; Moskalev *et al.*, 2011; Rein *et al.*, 1998; Rubatino *et al.*, 2015; Shen *et al.*, 2007; Voss *et al.*, 1998; Warnecke *et al.*, 1997; Warnecke *et al.*, 2002; Wojdacz and Hansen, 2006). To limit as much as possible a biased amplification, the T_m difference between the methylated PCR product and the unmethylated PCR product should be kept as low as possible.

As the difference between those two sequences, unmethylated and fully methylated, resides only in the CpG sites, with a variable base being either a C or a T, the GC content difference and associated bias depend on the CpG proportion. Thus, the number of CpG sites should be limited and should depend on the sequence length. Although many studies are associating the amplification bias with GC content and T_m difference, there is no mention of reducing the number of CpG, depending on the PCR product length, to limit this bias, in the literature.

Thirdly, the more CpG in the amplicon, the greater migration differences are between methylated and unmethylated templates during capillary electrophoresis, the higher is the risk of base

signals misalignment in sequencing results and biased estimation of methylation.

In consequence, the base composition of the amplicon is highly related to potential artifacts in the methylation quantification due to three main reasons: ① a high number of CpG in the amplicon increases the possibility of a biased PCR amplification towards methylation or unmethylation, ② an important amount of CpG in a long amplicon favors migration-induced misalignment of bases in sequencing results causing artifacts in methylation estimation, and ③ GC-rich sequences by containing a high quantity of converted C into T, have a reduced complexity of DNA which diminishes the prospective sites for specific primers and increases the frequency of T repeats, a cause of polymerase slippage and misalignment of bases derived artifacts. Therefore, it is preferable that the targeted sequence should not be located in GC-rich regions and should have a low CpG content, yet sufficient enough for the experiment relevance, to ensure a proper methylation estimation. Employing the BSP approach to study GC-rich regions, such as CpG islands, might be questionable — other methods might be more suited — and should be done with caution.

3.7 ADDITIONAL SEQUENCES

In several studies, the addition of a standard primer sequence at the 5' end of bisulfite primers allows the use of standard primers for sequencing to improve its effectiveness, as by adding cytosines bases the unbalanced nucleotide ratio is reduced and can lessen a potential cytosine signal over-scaling (Brisotto *et al.*, 2015). For example, in Voss *et al.* (1998), Chhibber and Schroeder (2008) and Brisotto *et al.* (2015), the authors used bisulfite primers tailed with M13 standard primer sequences.

4 CONCLUSION

To conclude about primer design for a bisulfite sequencing PCR application, it can be extremely difficult to respect all the reported recommendations in the literature. Some of the guidelines must sometimes be prioritized, and many possible primer pairs have to be compared and tested to achieve an unbiased and efficient PCR amplification of bisulfite-converted DNA.

KEY POINTS

- ➔ The guidelines for primer design are summarized in [Figure A1.1](#) and associated software in [Table A1.1](#).
- ➔ PCR on bisulfite-treated DNA comes with some issues mainly due to the loss in base diversity: low efficiency of amplification, diminished primer specificity, frequent undesired PCR product amplification, and frequent primer dimer formation.
- ➔ Even for methylation-independent PCR (MIP), primers can contain one CpG site, as far as possible from the 3' end.
- ➔ A minimum of converted cytosines must be included in the primer sequence to increase specificity towards converted DNA.
- ➔ Primers are generally around 25 bp long and PCR products around 200-300 bp long.
- ➔ Inter-primer melting temperature (T_m) difference and T_m difference between unmethylated and methylated sequences of PCR products must be as low as possible.
- ➔ A check for potential unintended PCR products is necessary to avoid contaminating PCR products.

↗ [Back to Table of Contents](#)

5 BIBLIOGRAPHY

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Anzaldi, L. J., Muñoz-Fernández, D., & Erill, I. (2012). BioWord: A sequence manipulation suite for Microsoft Word. *BMC Bioinformatics*, 13(1), 124–124. <https://doi.org/10.1186/1471-2105-13-124>
- Árányi, T., & Tusnády, G. E. (2007). PCR Primer Design. *Methods in Molecular Biology™*, 402, 385–402. https://doi.org/10.1007/978-1-59745-528-2_20
- Árányi, T., Váradi, A., Simon, I., & Tusnády, G. E. (2006). The BiSearch web server. *BMC Bioinformatics*, 7(1), 431–431. <https://doi.org/10.1186/1471-2105-7-431>
- Ashapkin, V. V., Kutueva, L. I., & Vanyushin, B. F. (2020). Clinical and Preclinical Models for Maximizing Healthspan, Methods and Protocols. *Methods in Molecular Biology*, 2138, 297–312. https://doi.org/10.1007/978-1-0716-0471-7_21
- Brisotto, G., Gennaro, A. d., Damiano, V., Armellin, M., Perin, T., Maestro, R., & Santarosa, M. (2015). An improved sequencing-based strategy to estimate locus-specific DNA methylation. *BMC Cancer*, 15(1), 639. <https://doi.org/10.1186/s12885-015-1646-6>
- Candiloro, I. L. M., Mikeska, T., & Dobrovic, A. (2017). Assessing alternative base substitutions at primer CpG sites to optimise unbiased PCR amplification of methylated sequences. *Clinical Epigenetics*, 9(1), 31. <https://doi.org/10.1186/s13148-017-0328-4>
- Chen, G. G., Gross, J. A., Lutz, P.-E., Vaillancourt, K., Maussion, G., Bramouille, A., Thérout, J.-F., Gardini, E. S., Ehlert, U., Bourret, G., Masurel, A., Lepage, P., Mechawar, N., Turecki, G., & Ernst, C. (2017a). Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics*, 18(1), 96. <https://doi.org/10.1186/s12864-017-3489-9>
- Chen, Y., Pal, B., Visvader, J. E., & Smyth, G. K. (2017b). Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Research*, 6, 2055. <https://doi.org/10.12688/f1000research.13196.2>
- Chhibber, A., & Schroeder, B. G. (2008). Single-molecule polymerase chain reaction reduces bias: Application to DNA methylation analysis by bisulfite sequencing. *Analytical Biochemistry*, 377(1), 46–54. <https://doi.org/10.1016/j.ab.2008.02.026>
- Clark, S. J., Harrison, J., Paul, C. L., & Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Research*, 22(15), 2990–2997. <https://doi.org/10.1093/nar/22.15.2990>

- Fazekas, A. J., Steeves, R., & Newmaster, S. G. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques*, 48(4), 277–285. <https://doi.org/10.2144/000113369>
- Fitzpatrick, D. R., Shirley, K. M., McDonald, L. E., Bielefeldt-Ohmann, H., Kay, G. F., & Kelso, A. (1998). Distinct Methylation of the Interferon γ (IFN- γ) and Interleukin 3 (IL-3) Genes in Newly Activated Primary CD8+ T Lymphocytes: Regional IFN- γ Promoter Demethylation and mRNA Expression Are Heritable in CD44^{high}CD8+ T Cells. *The Journal of Experimental Medicine*, 188(1), 103–117. <https://doi.org/10.1084/jem.188.1.103>
- Guldberg, P., Worm, J., & Grønbaek, K. (2002). Profiling DNA methylation by melting analysis. *Methods*, 27(2), 121–127. [https://doi.org/10.1016/s1046-2023\(02\)00063-4](https://doi.org/10.1016/s1046-2023(02)00063-4)
- Hernandez, H. G., Tse, M. Y., Pang, S. C., Arboleda, H., & Forero, D. A. (2013). Optimizing methodologies for PCR-based DNA methylation analysis. *BioTechniques*, 55(4), 181–197. <https://doi.org/10.2144/000114087>
- Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35(Web Server issue), W43–W46. <https://doi.org/10.1093/nar/gkm234>
- Kovacova, V., & Janousek, B. (2012). Bisprimer—A Program for the Design of Primers for Bisulfite-Based Genomic Sequencing of Both Plant and Mammalian DNA Samples. *Journal of Heredity*, 103(2), 308–312. <https://doi.org/10.1093/jhered/esr137>
- Kushwaha, G., Srivastava, G. P., & Xu, D. (2015). PCR Primer Design. *Methods in Molecular Biology*, 1275, 181–199. https://doi.org/10.1007/978-1-4939-2365-6_14
- Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3), 203–21. <https://doi.org/10.1093/oxfordjournals.molbev.a040442>
- Li, L.-C. (2007). PCR Primer Design. *Methods in Molecular Biology™*, 402, 370–383. https://doi.org/10.1007/978-1-59745-528-2_19
- Li, L.-C., & Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, 18(11), 1427–1431. <https://doi.org/10.1093/bioinformatics/18.11.1427>
- Li, R., Liang, F., Li, M., Zou, D., Sun, S., Zhao, Y., Zhao, W., Bao, Y., Xiao, J., & Zhang, Z. (2018). MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Research*, 46(Database issue), D288–D295. <https://doi.org/10.1093/nar/gkx1139>
- Li, Y., & Tollefsbol, T. O. (2011). Epigenetics Protocols. *Methods in Molecular Biology*, 791, 11–21. https://doi.org/10.1007/978-1-61779-316-5_2
- Lu, J., Johnston, A., Berichon, P., Ru, K.-I., Korbie, D., & Trau, M. (2017). PrimerSuite: A High-Throughput Web-Based Primer Design Program for Multiplex Bisulfite PCR. *Scientific Reports*, 7(1), 41328. <https://doi.org/10.1038/srep41328>
- Moskalev, E. A., Zavgorodnij, M. G., Majorova, S. P., Vorobjev, I. A., Jandaghi, P., Bure, I. V., & Hoheisel, J. D. (2011). Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Research*, 39(11), e77–e77. <https://doi.org/10.1093/nar/gkr213>
- Parrish, R. R., Day, J. J., & Lubin, F. D. (2012). Direct Bisulfite Sequencing for Examination of DNA Methylation with Gene and Nucleotide Resolution from Brain Tissues. *Current Protocols in Neuroscience*, 60(1), 7.24.1–7.24.12. <https://doi.org/10.1002/0471142301.ns0724s60>
- Poucke, M. V., Boulougouris, X., Spiegeleer, B. D., Burvenich, C., Duchateau, L., & Peelman, L. J. (2017). An optimized strategy for cloning-based locus-specific bisulfite sequencing PCR. *bioRxiv*, 239566. <https://doi.org/10.1101/239566>
- Rein, T., DePamphilis, M. L., & Zorbas, H. (1998). Identifying 5-methylcytosine and related modifications in DNA genomes. *Nucleic Acids Research*, 26(10), 2255–2264. <https://doi.org/10.1093/nar/26.10.2255>
- Robinson, J. T., Thorvaldsdóttir, H., Turner, D., & Mesirov, J. P. (2020). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *bioRxiv*, 2020.05.03.075499. <https://doi.org/10.1101/2020.05.03.075499>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for General Users and for Biologist Programmers. In {Krawetz, S. Misener and S.A}, H. P. Inc., Totowa, & NJ (Eds.). Humana Press. <https://doi.org/10.1385/1-59259-192-2-365>
- Rubatino, F., Carobin, N., Freitas, M., Oliveira, V. d., Pietra, R., Oliveira, P., Bosco, A., & Jehee, F. (2015). Manipulation of primer affinity improves high-resolution melting accuracy for imprinted genes. *Genetics and Molecular Research*, 14(3), 7864–7872. <https://doi.org/10.4238/2015.july.14.12>
- Rychlik, W. (1995). Selection of primers for polymerase chain reaction. *Molecular Biotechnology*, 3(2), 129–134. <https://doi.org/10.1007/bf02789108>
- Sehn, J. K. (2015). Clinical Genomics. *Section II: Bioinformatics*, 129–150. <https://doi.org/10.1016/b978-0-12-404748-8.00009-5>
- Shen, L., Guo, Y., Chen, X., Ahmed, S., & Issa, J.-P. J. (2007). Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. *BioTechniques*, 42(1), 48–58. <https://doi.org/10.2144/000112312>

- Shinde, D., Lai, Y., Sun, F., & Arnheim, N. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)_n and (A/T)_n microsatellites. *Nucleic Acids Research*, *31*(3), 974–980. <https://doi.org/10.1093/nar/gkg178>
- Tanaka, K., & Okamoto, A. (2007). Degradation of DNA by bisulfite treatment. *Bioorganic & Medicinal Chemistry Letters*, *17*(7), 1912–1915. <https://doi.org/10.1016/j.bmcl.2007.01.040>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Tost, J. (2010). DNA Methylation: An Introduction to the Biology and the Disease-Associated Changes of a Promising Biomarker. *Molecular Biotechnology*, *44*(1), 71–81. <https://doi.org/10.1007/s12033-009-9216-2>
- Tost, J., & Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. *Nature Protocols*, *2*(9), 2265–2275. <https://doi.org/10.1038/nprot.2007.314>
- Tusnady, G. E., Simon, I., Varadi, A., & Aranyi, T. (2005). BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Research*, *33*(1), e9–e9. <https://doi.org/10.1093/nar/gni012>
- Voss, K. O., Roos, K. P., Nonay, R. L., & Dovichi, N. J. (1998). Combating PCR Bias in Bisulfite-Based Cytosine Methylation Analysis. Betaine-Modified Cytosine Deamination PCR. *Analytical Chemistry*, *70*(18), 3818–3823. <https://doi.org/10.1021/ac980067t>
- Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., & Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Research*, *25*(21), 4422–4426. <https://doi.org/10.1093/nar/25.21.4422>
- Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., & Clark, S. J. (2002). Identification and resolution of artifacts in bisulfite sequencing. *Methods*, *27*(2), 101–107. [https://doi.org/10.1016/s1046-2023\(02\)00060-9](https://doi.org/10.1016/s1046-2023(02)00060-9)
- Wojdacz, T. K., & Hansen, L. L. (2006). Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. *BioTechniques*, *41*(3), 274–278. <https://doi.org/10.2144/000112240>
- Wojdacz, T. K., Hansen, L. L., & Dobrovic, A. (2008). A new approach to primer design for the control of PCR bias in methylation studies. *BMC Research Notes*, *1*(1), 54–54. <https://doi.org/10.1186/1756-0500-1-54>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134–134. <https://doi.org/10.1186/1471-2105-13-134>
- Zou, D., Sun, S., Li, R., Liu, J., Zhang, J., & Zhang, Z. (2015). MethBank: a database integrating next-generation sequencing single-base-resolution DNA methylation programming data. *Nucleic Acids Research*, *43*(Database issue), D54–D58. <https://doi.org/10.1093/nar/gku920>

APPENDIX 2

ROLE OF EXTRACELLULAR VESICLES IN RADIO-INDUCED DEDIFFERENTIATION INTO CANCER STEM CELLS

☰ CHAPTER CONTENTS

1	INTRODUCTION	319
1.1	EXTRACELLULAR VESICLES COMMUNICATION	319
1.2	EXTRACELLULAR VESICLES AND CANCER STEM CELLS	321
1.3	DEDIFFERENTIATION MEDIATED BY EXTRACELLULAR VESICLES	322
2	OBJECTIVE	323
3	MATERIAL AND METHODS	323
3.1	CELL LINES AND CULTURE CONDITIONS	323
3.2	EXTRACELLULAR VESICLES ISOLATION	323
3.3	EXTRACELLULAR VESICLES TREATMENT	324
3.4	RNA AND PROTEIN LEVELS	324
3.5	EXTRACELLULAR VESICLES UPTAKE	324
4	RESULTS	325
4.1	EXPERIMENTAL STRATEGY	325
4.2	CHARACTERIZATION OF THE RADIO-INDUCED EXTRACELLULAR VESICLES CONTENT	326
4.3	EFFECT OF RADIO-INDUCED EXTRACELLULAR VESICLES ON UNIRRADIATED CELLS	328
5	CONCLUSION	331
	📌 KEY POINTS	331
6	BIBLIOGRAPHY	331

[^ Back to Table of Contents](#)

1 INTRODUCTION

1.1 EXTRACELLULAR VESICLES COMMUNICATION

Extracellular vesicles (EVs) are secreted by cells and defined as a heterogeneous population of cell-derived membrane vesicles. They mediate intercellular communication as they transfer a molecular cargo encapsulated in a double-layered membrane from one cell to another (Figure A2.1). Their cargo is composed of several bioactive molecules including proteins, lipids, and nucleic acids (e.g. mRNAs, miRNAs).

Three main classes of EVs can be distinguished depending on their diameter and biogenesis: **exosomes**, **microvesicles**, and **apoptotic bodies** (Figure A2.1) (Bebelman *et al.*, 2018; Doyle and

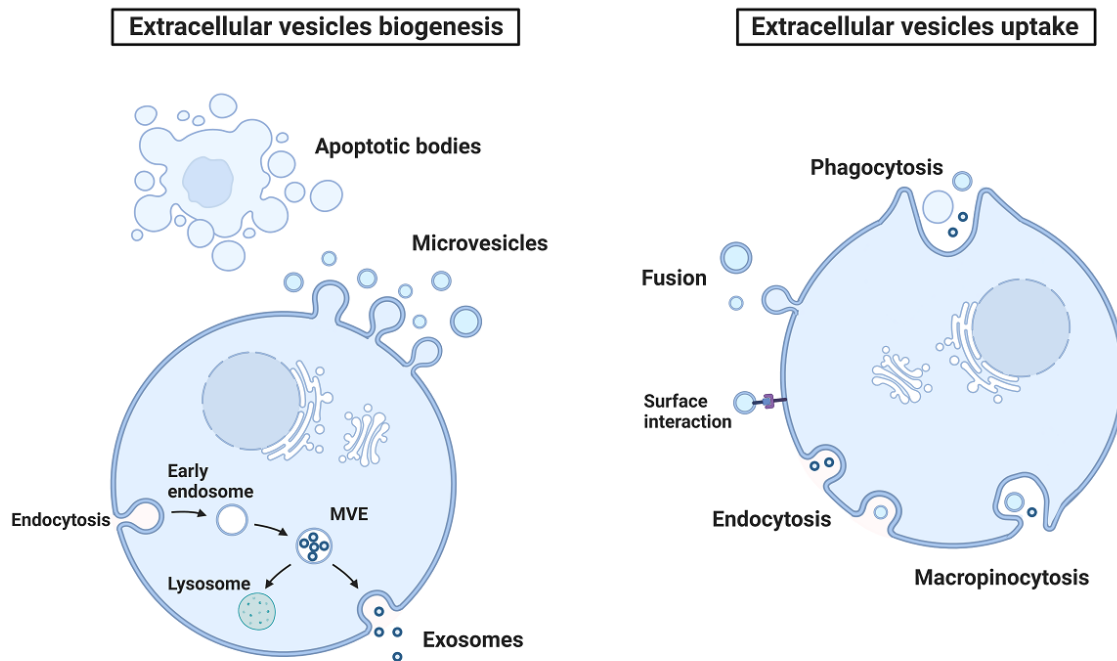


FIGURE A2.1 EXTRACELLULAR VESICLES COMMUNICATION: BIOGENESIS AND UPTAKE. Exosomes, microvesicles, and apoptotic bodies are secreted by cells and act as messengers in the extracellular space transporting molecular cargos to recipient cells that can uptake extracellular vesicles *via* fusion, endocytosis, macropinocytosis, and phagocytosis processes. Created with BioRender.com

Wang, 2019). **Exosomes** have a diameter generally between 30 to 150 nm and are formed in endosomes within the cell. By inward budding of the endosome membrane, intraluminal vesicles (ILVs) are formed in the multivesicular endosomes (MVEs) also known as multivesicular bodies (MVBs), that can dock and fuse with the plasma membrane to release the enclosed vesicles, then called exosomes, in the extracellular compartment. **Microvesicles** are generally ranging from 150 to 1000 nm in diameter and are created by the outward budding of the cell plasma membrane to be released into the extracellular environment. **Apoptotic bodies**, or apoptosomes, have a diameter in general around 500 to 5000 nm even if smaller ones (50-500 nm) can also be found, and originate from dying cells as the cellular content disintegrates into distinct membrane-enclosed vesicles.

The EVs uptake is first mediated by the surface interaction between vesicles surface ligands and cell plasma membrane-specific receptors and then the recipient cell can internalize the EVs through various processes such as membrane fusion, endocytosis, macropinocytosis, and phagocytosis (Figure A2.1). The EVs content is therefore released in the cytoplasm or endosomes and can regulate many cell signaling pathways.

In cancers, EVs participates in tumor progression, metastasis and therapeutic resistance (Becker *et al.*, 2016; Li and Nabet, 2019; Maacha *et al.*, 2019; Namee and O'Driscoll, 2018; Scholl *et al.*, 2020). Indeed, oncogenic and regulatory molecules have been found in EVs cargoes, such as

epidermal growth factor (EGF) receptors, activating oncogenic pathways in recipient cells and leading to accelerated growth and tumor progression (Al-Nedawi *et al.*, 2008). They have a role in the therapeutic resistance as EVs can mediate the sequestration or export of cytotoxic drugs to reduce their concentration, and activation of survival signaling pathways through intercellular communication to confer resistance properties to recipient cells (Chapuy *et al.*, 2008; Ifergan *et al.*, 2005; Maacha *et al.*, 2019; Safaei *et al.*, 2005).

Following radiotherapy, secreted EVs participate in the radiation-induced bystander effect, meaning that even unirradiated cells can be affected by radiotherapy effects through the uptake of EVs derived from irradiated neighboring ones (Jabbari *et al.*, 2020; Jella *et al.*, 2014; Szatmári *et al.*, 2017; Xu *et al.*, 2016). Indeed, EVs can transmit the oxidative stress generated by radiotherapy in irradiated cells to recipient cells, expanding the ROS-mediated DNA damage to unirradiated cells. Moreover, the radiation treatment induces changes in the secreted EVs quantity and content, which can confer radiation resistance properties to recipient cells (Al-Mayah *et al.*, 2012; Jelonek *et al.*, 2016). The transmission of bystander information *via* EV communication has been reported to promote cell survival after radiotherapy (Mutschelknaus *et al.*, 2016).

1.2 EXTRACELLULAR VESICLES AND CANCER STEM CELLS

The EVs produced by CSC themselves have been reported to promote proliferation, migration, angiogenesis, and metastasis (Andrés *et al.*, 2020; Nawaz, 2017; Su *et al.*, 2021; Wang *et al.*, 2020; Wang *et al.*, 2019). For example, the miR-19b-3p is transported by CSC-derived exosomes and promotes the EMT *via* the repression of phosphatase and tensin homolog (PTEN) expression in the clear cell renal cell carcinoma (CCRCC) (Wang *et al.*, 2019). In hepatocellular carcinoma, exosomes released by CSCs induce the expression of NANOG in non-CSCs and enhance their drug resistance (Huang *et al.*, 2021).

The radio-induced EVs and chemo-induced EVs promote stem cell characteristics, amplifying the therapeutic resistance (Chung *et al.*, 2021; Kyjacova *et al.*, 2015; Meldolesi, 2022; Ramakrishnan *et al.*, 2020; Shen *et al.*, 2019; Yang *et al.*, 2021). For instance, Chung *et al.* (2021) showed by using two bladder cancer sub-lines, one with CSC-like properties and the other with a non-CSC phenotype, that chemo-induced EVs from the non-CSC cell line enhanced the sphere formation and tumorigenicity of the CSC-like cell line, highlighting the role of EVs from neighboring non-CSCs for maintenance of stemness characteristics and expansion of the CSCs in response to chemotherapy. In breast cancer, Shen *et al.* (2019) observed that chemo-induced EVs enriched in miRNAs promote the CSC phenotype *via* the down-regulation of the master transcription factor ONECUT2, which is generally associated with a differentiated phenotype. Concerning radiation treatments, Ramakrishnan *et al.* (2020) reported in glioblastoma that specific miRNAs can be exported outside the cell by vesicle encapsulation and released in response to radiotherapy. The release of the miR-603 cause a derepression of the insulin-like growth factor 1 (IGF1) and its receptor, promoting the CSC state and

radioresistance of the cell *via* the up-regulation of DNA repair.

Additionally, EVs derived from stromal cells also participate in the promotion of stem cell features and in the radio- or chemo-resistance of cancer cells. CAFs-derived EVs contribute to the maintenance and expansion of colorectal CSCs through activation of stemness related pathways, such as the Wnt/ β -catenin signaling or TGF- β signaling, mediated by EV-transported miRNAs (e.g. miR-92a-3p) or lncRNAs (e.g. H19) (Conigliaro *et al.*, 2015; Hu *et al.*, 2019a; Hu *et al.*, 2015; Liu *et al.*, 2020; Ren *et al.*, 2018). In breast cancer, EVs from stromal cells containing three miRNAs (miR-21, miR-378e, and miR-143) increase the mammospheres formation and promote the EMT and stemness phenotypes (Donnarumma *et al.*, 2017), and they can activate the STAT1 and Notch signaling promoting stemness and resistance to therapy (Boelens *et al.*, 2014). In glioblastoma, The macrophage-derived EVs uptake by CSCs results in elevated CSC marker expression, promotion of proneural-to-mesenchymal transition and increased radioresistance (Zhang *et al.*, 2020).

1.3 DEDIFFERENTIATION MEDIATED BY EXTRACELLULAR VESICLES

The interconversion between CSCs and non-CSCs is mediated by EVs. Indeed, in their cargo, stemness regulatory molecules can induce a phenotypic switch of recipient cells, to maintain the balance between CSCs and non-CSCs or expand the CSC pool in response to therapy for instance (Sun *et al.*, 2018; Xu *et al.*, 2018).

First, to study the effect of CSC-derived exosomes on non-CSCs, Sun *et al.* (2020) used three glioblastoma non-CSC cell lines (WJ1, U251, and U87) treated with exosomes from a CSC-like cell line (WJ2) cultured in a stem cell medium and showing high expression of CSC markers. Although the CSC and non-CSC populations were not purified using CSC markers, they reported an increase in tumorigenicity in the treated cells and enrichment of CSC-derived exosomes in Notch1 proteins. As the inhibition of Notch1 decreased the stemness features induced by CSC-exosomes in non-CSC cell lines, it revealed that the Notch signaling contributes to the exosome-mediated enhanced stemness of glioblastoma cells.

In colorectal cancer, Hu *et al.* (2019b) demonstrated the induced dedifferentiation of purified non-CSCs by CAFs-derived EVs during chemotherapy treatment. The non-CSC population treated with fibroblast-derived conditioned media showed higher sphere formation and higher CSC percentage upon chemotherapy treatment. Furthermore, the fibroblast-derived exosomes coupled with chemotherapy-induced more sphere generation in the non-CSC population compared to control cells, while the exosome-depleted conditioned media treatment induced less sphere generation, indicating that the dedifferentiation is mediated by exosomes. They also found the presence of several Wnt ligands in the fibroblast-secreted vesicles, activating the Wnt/ β -catenin pathway in recipient cells. The overexpression of Wnt3a in CAFs revealed that CAFs-derived EVs increases the nuclear β -catenin levels in treated non-CSCs and enhances their tumorigenicity *in vivo*.

2 OBJECTIVE

The study presented here aims to know if radio-induced EVs are implicated in the non-CSC-to-CSC conversion. Indeed, after irradiation, the EVs could transport molecules such as proteins or mRNA that act as messengers in recipient cells to unlock the phenotypic switch toward a stem state. We hypothesize here that mRNAs of genes associated with phenotypic conversion, either with the CSC plasticity or with the EMT, may be carried in the radio-induced EVs to induce the dedifferentiation of surrounding recipient non-CSCs into CSCs.

3 MATERIAL AND METHODS

This section completes the main material and methods, at page 153, with techniques and protocols specifically used for the study of the extracellular vesicles involvement in radio-induced dedifferentiation.

3.1 CELL LINES AND CULTURE CONDITIONS

Additionally to the previously used SUM159PT cells, the MDA-MB-231, MCF7, and T47D (ATCC) breast cancer cell lines were used to study the extracellular vesicle communication throughout the radio-induced dedifferentiation of cancer cells. Their main characteristics are depicted in [Table A2.1](#). These cell lines were cultivated in the culture media described in [Table A2.2](#). The Ham's F-12 Nutrient Mixture (F-12), Minimum Essential Medium (MEM), Roswell Park Memorial Institute 1640 medium with GlutaMAX (RPMI GlutaMAX), Non-Essential Amino Acids (NEAA), HEPES buffer, penicillin, and streptomycin are provided by Gibco; the Fetal Bovine Serum (FBS) by HyClone; the insulin and hydrocortisone by Sigma-Aldrich and the ZellShield by Biovalley.

	SUM159PT	MDA-MB-231	MCF7	T47D
Supplier	Asterand	ATCC	ATCC	ATCC
Molecular type	Triple negative	Triple negative	Luminal A	Luminal A
Morphology	Epithelial	Epithelial	Epithelial	Epithelial
Tumor type	Anaplastic carcinoma	Adenocarcinoma	Adenocarcinoma	Ductal carcinoma
Sampling origin	Primary tumor	Pleural effusion	Pleural effusion	Pleural effusion

TABLE A2.1 CHARACTERISTICS OF BREAST CANCER CELL LINES.

3.2 EXTRACELLULAR VESICLES ISOLATION

Donor cells are seeded in 8 Petri plates (100 mm diameter) at 150,000 cells per plate one day before irradiation. Before irradiation, the medium is replaced by an EV-free medium (the fetal bovine serum is ultracentrifuged to remove EV prior to being added to the medium). The conditioned

	SUM159PT	MDA-MB-231	MCF7	T47D
Base medium	F-12	MEM	MEM	RPMI GlutaMAX
FBS	5%	10%	10%	10%
Insulin	5 µg/mL	-	10 µg/mL	10 µg/mL
Hydrocortison	1 µg/mL	-	-	-
NEAA	-	1%	1%	1%
HEPES	10 nM	-	-	-
Penicillin	-	-	40 units/mL	40 units/mL
Streptomycin	-	-	40 µg/mL	40 µg/mL
ZellShield	1%	1%	-	-

TABLE A2.2 COMPONENTS OF BREAST CANCER CELL LINES CULTURE MEDIA.

medium from the 8 Petri plates (80 mL of medium) of donor cells is collected five days after irradiation. It is first centrifuged at 500 g for 5 min at 4 °C to remove dead cells and at 3,000 g for 20 min at 4 °C to remove apoptotic bodies. The medium is then concentrated using the Amicon Ultra-15 100 kDa tubes with filters (Millipore) for a > 100 kDa filtration by centrifugation at 4,000 rpm for 40 min at 4 °C. The concentrated medium is ultracentrifuged using the Optima XPN-80 Beckman Coulter ultracentrifuge, with a 70.1Ti rotor, at 20,000 rpm (approx 30,000 g) for 2h30 at 4 °C. The pellet containing the purified EVs is either resuspended in 100 µL of PBS (phosphate-buffered saline) and stored at 4 °C for a maximum of 2 days before treatment, or resuspended in 200 µL of lysis buffer for RNA extraction.

3.3 EXTRACELLULAR VESICLES TREATMENT

The 100 µL of purified EVs diluted in PBS is used to treat 2 wells of a 6-well plate, each well seeded with 100,000 recipient cells the day before.

3.4 RNA AND PROTEIN LEVELS

The same protocols described in the main matter were used for RNA and protein levels evaluation using qPCR and Western blot techniques. Primers and antibodies used are listed in [Table A2.3](#) and [Table A2.4](#).

3.5 EXTRACELLULAR VESICLES UPTAKE

Membranes of donor cells (SUM159PT cells) were stained with the Vybrant DiI label from Invitrogen according to the manufacturer's instructions before irradiation at 0, 4, or 8 Gy. The cell medium is collected 5 days after irradiation and centrifuged at 1,000 rpm to remove dead cells. Sorted ALDH⁺ recipient cells were treated with the conditioned medium containing the unlabelled or DiI-labelled EVs (half of conditioned medium and half of fresh medium). The DiI-label EVs uptake

TARGET GENE	SEQUENCE FORWARD	SEQUENCE REVERSE
<i>B2M</i>	5'-TCGCGCTACTCTCTCTTT-3'	5'-CAAGTCTGAATGCTCCACTT-3'
<i>RPLP0</i>	5'-GCGACCTGGAAGTCCAACATA-3'	5'-TGTCTGCTCCCACAATGAAG-3'
<i>POU5F1</i>	5'-GAAGGATGTGGTCCGAGTGT-3'	5'-GTGAAGTGAGGGCTCCCATA-3'
<i>SOX2</i>	5'-AACCCCAAGATGCACACCTC-3'	5'-CGGGGCCGGTATTATAATC-3'
<i>NANOG</i>	5'-GTGATTTGTGGCCTGAAGA-3'	5'-ACACAGCTGGGTGGAAGAGA-3'
<i>ALDH1A1</i>	5'-GTTGAGCGGGCTAAGTAT-3'	5'-CCCCTCTCAATGAGGTCAAG-3'
<i>C-MYC</i>	5'-TTGTCTCCGTCGGATTCTCTG-3'	5'-TCTTCTTGTTCCTCCTCAGAGTCG-3'
<i>KLF4</i>	5'-TACCAAGAGCTCATGCCACC-3'	5'-GGTGTGCCTTGAGATGGGAA-3'
<i>CD44</i>	5'-AGACATCTACCCAGCAAC-3'	5'-CGTTGAGTCCACTTGGCTTTC-3'
<i>CDH1</i>	5'-GTCAGTTCAGACTCCAGCCC-3'	5'-AAATCACTCTGCCCAGGACG-3'
<i>VIM</i>	5'-TCTACGAGGAGGAGATGCGG-3'	5'-GGTCAAGACGTGCCAGAGAC-3'
<i>SNAI1</i>	5'-ACCACTATGCCGCGCTCTT-3'	5'-GGTCGTAGGGCTGCTGGAA-3'
<i>SNAI2</i>	5'-TGTTGCAGTGAGGGCAAGAA-3'	5'-GACCCTGGTTGCTTCAAGGA-3'
<i>ZEB1</i>	5'-CTGATCCCCAGGTGGCATA-3'	5'-GGGCGGTGTAGAATCAGAGT-3'
<i>TWIST2</i>	5'-CGACGAGATGGACAATAAGATGAC-3'	5'-CAGGTTTCAGAAGTTACAGACTCG-3'

TABLE A2.3 LIST OF SPECIFIC PRIMERS FOR RT-QPCR.

TARGET PROTEIN	SUPPLIER	REFERENCE	SPECIE	DILUTION
Nanog	Cell Signaling	#3580	Rabbit	1:1000
Oct4	Cell Signaling	#2750	Rabbit	1:1000
SOX2	Cell Signaling	#2748	Rabbit	1:1000
ALDH1	Abcam	#ab52492	Rabbit	1:1000
β -actin	Sigma-Aldrich	#A2066	Rabbit	1:500
Rabbit IgG	Jackson ImmunoResearch	#711-035-152	Donkey	1:5000

TABLE A2.4 LIST OF ANTIBODIES USED FOR WESTERN BLOT.

by recipient cells was assessed 5 days later using a fluorescence microscope (Nikon Eclipse Ti, 10X objective).

4 RESULTS

4.1 EXPERIMENTAL STRATEGY

The [Figure A2.2](#) illustrates the strategy used to study both the radio-induced EV content and its effect on the dedifferentiation of non-CSCs into CSCs.

First, to characterize the molecular content of radio-induced EVs, donor cells are seeded and

irradiated the next day at 0 Gy or 8 Gy (Figure A2.2). The conditioned media from donor cells are collected 5 days later and EVs are purified by ultracentrifugation. The mRNA and protein content of EVs is analyzed by RT-qPCR and Western blot.

Then, to analyze the generation of CSCs by radio-induced EVs, recipient cells are treated with the freshly purified EVs from irradiated or unirradiated cells, one day after being seeded (Figure A2.2). In parallel, untreated control cells are irradiated at 0 Gy or 8 Gy on the same day. The CSC percentage and mRNA expression of prospective genes is assessed 5 days later.

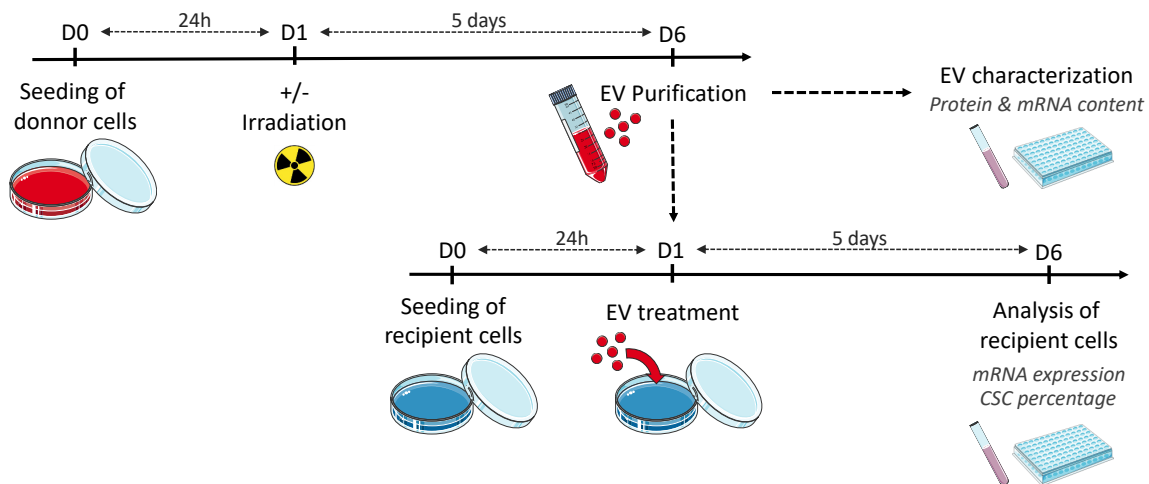


FIGURE A2.2 EXPERIMENTAL STRATEGY TO STUDY THE EFFECT OF IRRADIATION ON EXTRACELLULAR VESICLE COMMUNICATION AND ITS ROLE IN THE DEDIFFERENTIATION OF CANCER CELLS.

4.2 CHARACTERIZATION OF THE RADIO-INDUCED EXTRACELLULAR VESICLES CONTENT

EVs from SUM159PT cells were previously characterized in the laboratory, they express the HSP90, Flotilin-1, TSG101, and Caveolin-1 EV markers and their production is increased after irradiation at 8 Gy (data not shown).

First, mRNA levels of 10 genes associated with stemness or EMT: OCT4, C-MYC, KLF4, CD44, VIM (vimentin), SNAI1, SNAI2, ZEB and TWIST2, have been measured in SUM159PT and MDA-MB-231 breast cancer cells and in their secreted EV, after a 0 Gy or 8 Gy radiation treatment (Figure A2.3A and B). Both in SUM159PT and MDA-MB-231 cells, SNAI1 and SNAI2 genes tend to be upregulated after radiation exposure (Figure A2.3A). In SUM159PT cells, SNAI1 and SNAI2 mRNA levels are increased 7.48 fold and 4.99 fold after irradiation, respectively. In MDA-MB-231 cells, SNAI1 and SNAI2 mRNA levels are increased by 3.05 ± 1.36 fold and 3.29 ± 1.43 fold after irradiation, respectively. For SUM159PT cells, CD44, SNAI1, and SNAI2 mRNAs tends to be enriched in the radio-induced EVs, with an increase of 3.22 ± 0.41 , 3.73 ± 2.04 , and 3.31 ± 0.42 fold after irradiation, respectively (Figure A2.3B). For MDA-MB-231 cells, C-MYC, SNAI1, and SNAI2 mRNAs are significantly enriched in the radio-induced EVs, with an increase of 2.58 ± 0.51 ($p=0.041$), $3.18 \pm$

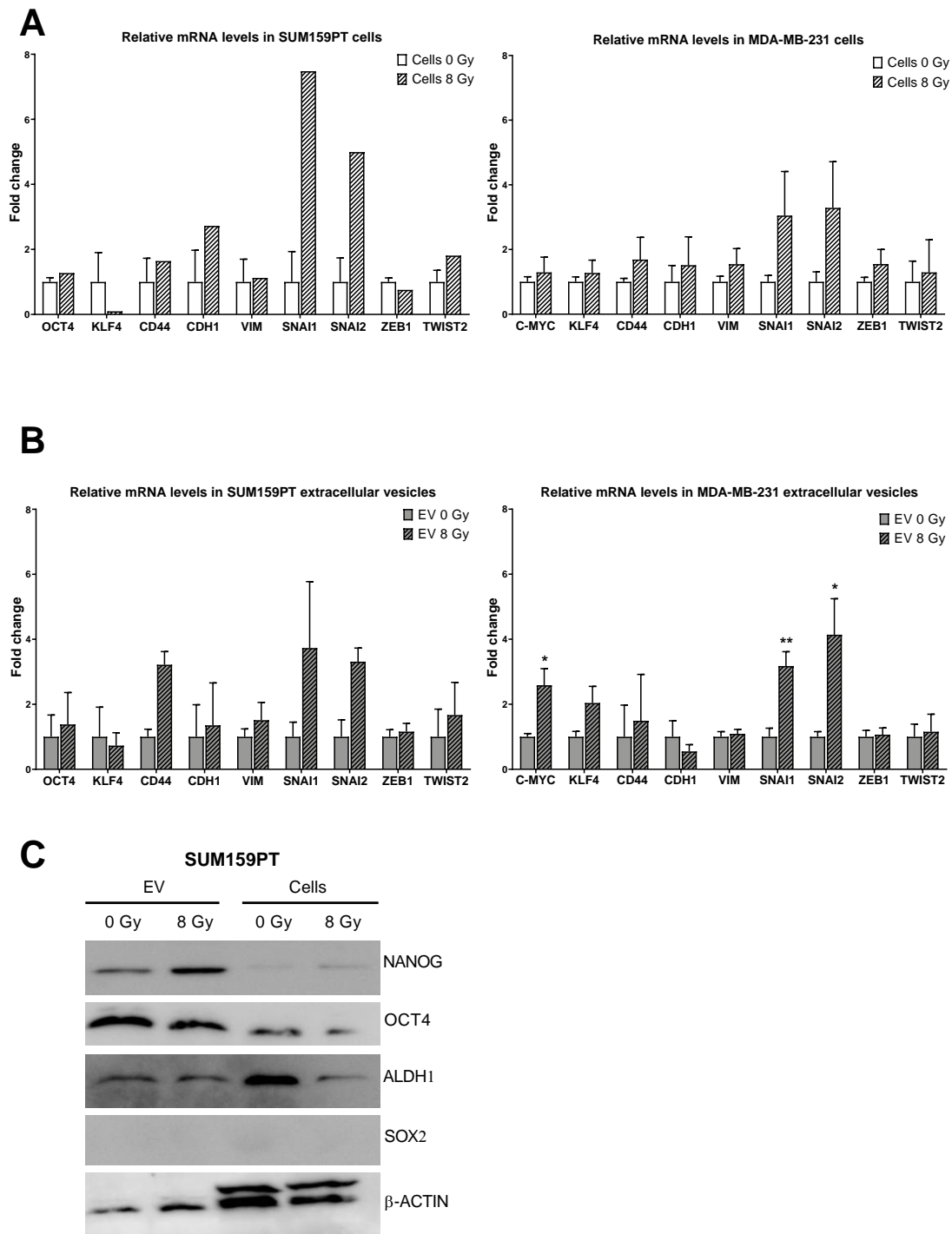


FIGURE A2.3 CHARACTERIZATION OF THE EXTRACELLULAR VESICLE CARGO AFTER RADIOTHERAPY TREATMENT. **A.** RNA levels of stemness and EMT associated genes after 0 Gy or 8 Gy irradiation in SUM159PT (left, $n=2$ for 0 Gy and $n=1$ for 8 Gy) and MDA-MB-231 (right, $n=3$) cell lines. **B.** RNA levels of stemness and EMT-associated genes after 0 Gy or 8 Gy irradiation in EVs from the SUM159PT (left, $n=2$) and MDA-MB-231 (right, $n=5$) cell lines. Error bars represent means \pm standard error of the mean. * Student's T-test p -value ≤ 0.05 ; ** Student's T-test p -value ≤ 0.01 . **C.** Protein levels of stemness-associated genes after 0 Gy or 8 Gy irradiation in both SUM159PT cells and their secreted EVs.

0.44 ($p=0.002$), and 4.14 ± 1.11 ($p=0.034$) fold after irradiation, respectively. These results indicate that SNAI1 and SNAI2 mRNAs are enriched in secreted EVs after radiation exposure, possibly due to their increased expression, both in SUM159PT and MDA-MB-231 cells. Additionally, C-MYC mRNA is also enriched in EVs from MDA-MB-231 cells.

Then, protein levels of NANOG, OCT4, ALDH1, and SOX2 were measured both in SUM159PT cells and EVs (Figure A2.3C). Interestingly, the NANOG protein is enriched in EVs after radiation exposure. The OCT4 and ALDH1 protein levels are similar in EVs between unirradiated and irradiated conditions, while in cells, the ALDH1 protein level is higher in unirradiated cells compared to irradiated ones, which is surprising and inconsistent with the literature (Bidan *et al.*, 2019). The SOX2 protein levels were too low to be detected both in cells and EVs.

Altogether, these results show that C-MYC, SNAI1, and SNAI2 mRNAs and NANOG proteins are enriched in radio-induced EVs from breast cancer cells, and this specific cargo could affect the recipient cells in response to radiotherapy.

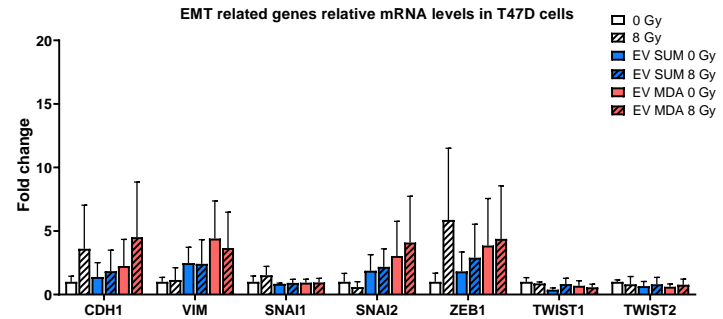
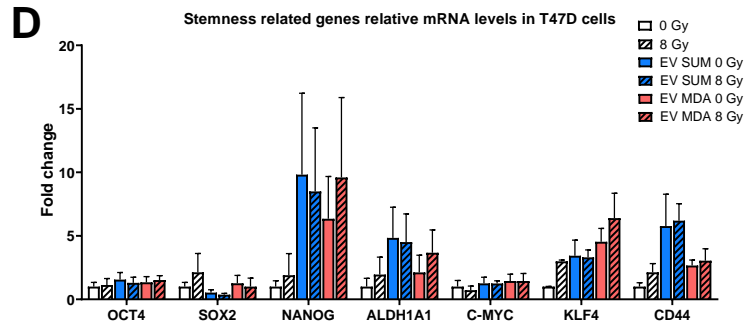
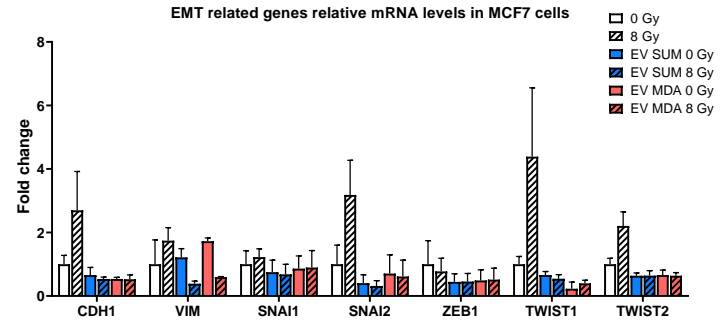
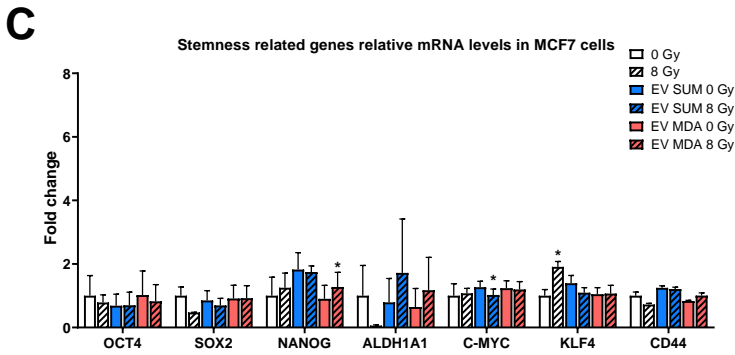
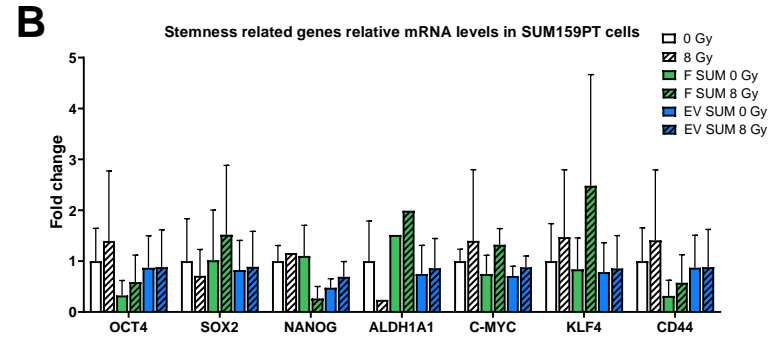
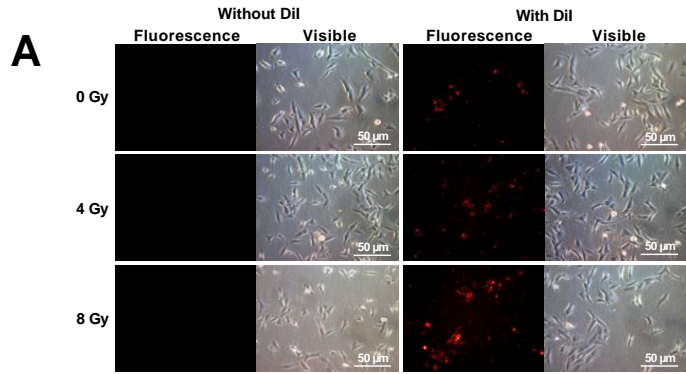
4.3 EFFECT OF RADIO-INDUCED EXTRACELLULAR VESICLES ON UNIRRADIATED CELLS

As we previously saw that radio-induced EVs have a specific cargo, it could affect their uptake by recipient cells following radiations exposure, and an increased uptake could be a cause of phenotype conversion.

So, a first experiment was conducted to know if the EV uptake of the recipient differs depending if donor cells were irradiated or not (Figure A2.4A). The conditioned medium from irradiated Dil-labeled donor cells at 0, 4, or 8 Gy was collected and added to the media of unirradiated recipient sorted non-CSCs (ALDH⁻ cells). The EV uptake is visible by microscopy as the Dil-labeling fluorescence from donor cells is transferred to recipient cells through EVs. Interestingly, the EV uptake seems slightly increased after the 4 Gy irradiation of donor cells and even more increased after the 8 Gy irradiation of donor cells.

To get insights into the radio-induced EVs contribution to the cell phenotypic plasticity, unirradiated cells were treated with EVs from irradiated cells and changes in expression of stemness and EMT-related genes were evaluated. To mimic the intratumoral heterogeneity, several combinations of cell lines as donor and recipient cells were tested.

In Figure A2.4B, SUM159PT cells were treated with EVs from SUM159PT cells. As controls, the non-EV part of the media (filtrate, F) was also assessed. For all the stemness-related genes tested, OCT4, SOX2, NANOG, ALDH1A1, C-MYC, KLF4, and CD44, none showed a significant change in expression in the recipient cells between the two EV treated conditions either coming from irradiated or unirradiated cells.



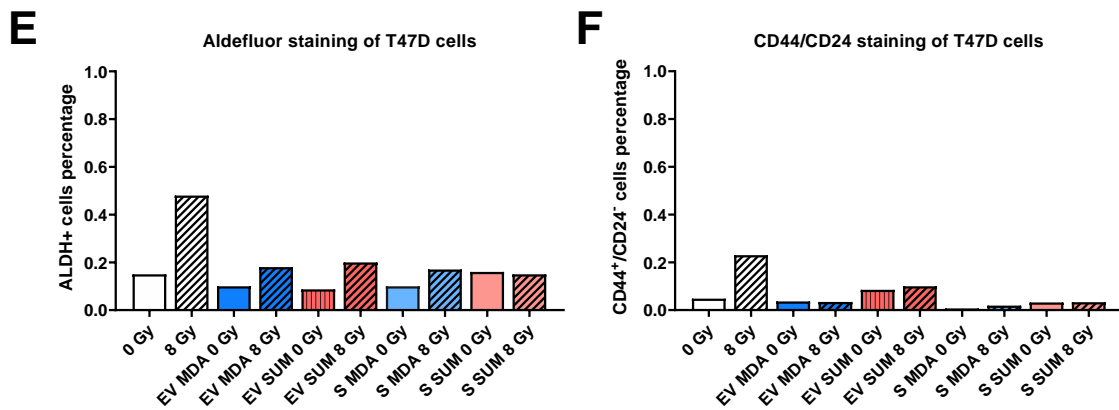


FIGURE A2.4 EFFECT OF RADIO-INDUCED EXTRACELLULAR VESICLES ON UNIRRADIATED CELLS GENE EXPRESSION AND PERCENTAGE OF CANCER STEM CELLS. **A.** Fluorescence images of EV uptake differences between EVs irradiated at 0, 4 or 8 Gy. The donor cell membranes were labeled with Dil before irradiation. Labeled-EVs were collected 5 days after irradiation and used to treat unirradiated cells. The EV uptake was assessed by fluorescence microscopy 5 days after the EV treatment. **B.** Relative mRNA levels of genes associated with stemness, in SUM159PT cells treated with either 0 or 8 Gy radiations, filtered conditioned media (filtrate, F) without EVs from irradiated or unirradiated cells, or EVs from irradiated or unirradiated cells (n=2). **C.** Relative mRNA levels of genes associated with stemness or EMT, in MCF7 cells treated with either 0 or 8 Gy radiations, or EVs from irradiated or unirradiated SUM159PT cells, or EVs from irradiated or unirradiated MDA-MB-231 cells (n=3). **D.** Relative mRNA levels of genes associated with stemness or EMT, in TD47 cells treated with either 0 or 8 Gy radiations, or EVs from irradiated or unirradiated SUM159PT cells, or EVs from irradiated or unirradiated MDA-MB-231 cells (n=3). Error bars represent means \pm standard error of the mean. * Student's T-test p-value \leq 0.05 **E. and F.** CSC percentages based on Aldefluor staining (ALDH⁺ cells) and CD44/CD24 staining (CD44⁺/CD24⁻ cells) in TD47 cells treated with either 0 or 8 Gy radiations, EVs from irradiated or unirradiated SUM159PT cells, EVs from irradiated or unirradiated MDA-MB-231 cells, conditioned media depleted in EVs (supernatant, S) from irradiated or unirradiated SUM159PT cells, conditioned media depleted in EVs (supernatant, S) from irradiated or unirradiated MDA-MB-231 cells (n=1).

In **Figure A2.4C**, MCF7 cells were treated with EVs from either SUM159PT or MDA-MB-231 cells. The expression of NANOG is increased after the 8 Gy EV from MDA-MB-231 treatment (1.28 ± 0.46 fold) compared to the 0 Gy EV from MDA-MB-231 treatment (0.90 ± 0.42 , $p=0.021$) while the expression of C-MYC is reduced after the 8 Gy EV from SUM159PT treatment (1.03 ± 0.18) compared to the 0 Gy EV from SUM159PT treatment (1.27 ± 0.19 , $p=0.030$). Except for these two genes, no other significant gene expression changes were observed.

In **Figure A2.4C**, T47D cells were treated with EVs from either SUM159PT or MDA-MB-231 cells but no significant differences in expression levels of these stemness and EMT-associated genes were found.

Lastly, the effect of the radio-induced EVs on the generation of CSCs was assessed by flow cytometry using both the CD44⁺/CD24⁻ population surface markers and the ALDH⁺ activity marker to measure the CSC proportion (**Figure A2.4E and F**). CSC proportions were surprisingly low, even in

the control conditions 0 Gy and 8 Gy in which a radio-induced increase of CSCs is noticeable. And, no variations in CSC percentages are observed in EV treatment conditions, so the experiment was not pursued.

5 CONCLUSION

Although some differences in EV cargoes were found between EVs from irradiated cells and EVs from unirradiated cells, the radio-induced EVs did not seem to strongly affect the gene expression programs of recipient cells nor the generation of CSCs. Therefore, these results are not conclusive about the role of radio-induced EVs in the enrichment in CSCs.

KEY POINTS

- ➔ Radio-induced EVs have been shown to promote stem cell properties and radioresistance through the transportation of a specific molecular content between cells (Kyjacova *et al.*, 2015; Ramakrishnan *et al.*, 2020).
- ➔ The EV communication could contribute to the acquisition of stem cell features leading to the dedifferentiation of non-CSCs into CSCs (Hu *et al.*, 2019b; Sun *et al.*, 2018; Sun *et al.*, 2020; Xu *et al.*, 2018).
- ➔ The characterization of EVs from irradiated SUM159PT cells revealed a specific molecular cargo enriched in C-MYC, SNAI1, and SNAI2 mRNAs and in NANOG proteins.
- ➔ MCF7 cells treated with EVs from irradiated MDA-MB-231 showed an increased expression of NANOG, while MCF7 cells treated with EVs from irradiated SUM159PT showed a reduced expression of C-MYC.
- ➔ No changes in CSC proportions were found in T47D cells after EV treatments from unirradiated or irradiated cells, hence the role of radio-induced EVs in the CSC enrichment could not have been determined by this approach.

↗ [Back to Table of Contents](#)

6 BIBLIOGRAPHY

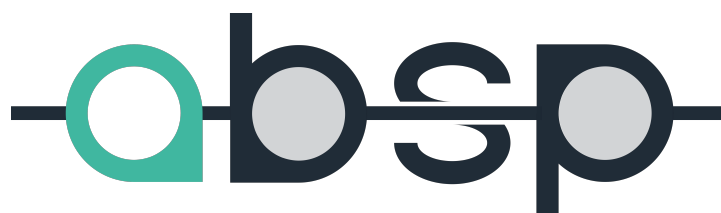
- Al-Mayah, A. H. J., Irons, S. L., Pink, R. C., Carter, D. R. F., & Kadhim, M. A. (2012). Possible Role of Exosomes Containing RNA in Mediating Nontargeted Effect of Ionizing Radiation. *Radiation Research*, 177(5), 539–545. <https://doi.org/10.1667/rr2868.1>
- Al-Nedawi, K., Meehan, B., Micallef, J., Lhotak, V., May, L., Guha, A., & Rak, J. (2008). Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nature Cell Biology*, 10(5), 619–624. <https://doi.org/10.1038/ncb1725>
- Andrés, J. L. d., Griñán-Lisón, C., Jiménez, G., & Marchal, J. A. (2020). Cancer stem cell secretome in the tumor microenvironment: a key point for an effective personalized cancer treatment. *Journal of Hematology & Oncology*, 13(1), 136. <https://doi.org/10.1186/s13045-020-00966-3>
- Bebelman, M. P., Smit, M. J., Pegtel, D. M., & Baglio, S. R. (2018). Biogenesis and function of extracellular vesicles in cancer. *Pharmacology & Therapeutics*, 188, 1–11. <https://doi.org/10.1016/j.pharmthera.2018.02.013>
- Becker, A., Thakur, B. K., Weiss, J. M., Kim, H. S., Peinado, H., & Lyden, D. (2016). Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. *Cancer Cell*, 30(6), 836–848. <https://doi.org/10.1016/j.ccell.2016.10.009>
- Bidan, N., Bailleul-Dubois, J., Duval, J., Winter, M., Denoulet, M., Hannebicque, K., El-Sayed, I. Y., Ginestier, C., Forissier, V., Charafe-Jauffret, E., Macario, M., Matsunaga, Y. T., Meignan, S., Anquez, F., Julien, S., Bonnefond, A., Derhourhi, M., Bourhis, X. L., & Lagadec, C. (2019). Transcriptomic Analysis of Breast Cancer Stem Cells and Development of

- a pALDH1A1:mNeptune Reporter System for Live Tracking. *PROTEOMICS*, 19(21-22), 1800454. <https://doi.org/10.1002/pmic.201800454>
- Boelens, M. C., Wu, T. J., Nabet, B. Y., Xu, B., Qiu, Y., Yoon, T., Azzam, D. J., Twyman-Saint Victor, C., Wiemann, B. Z., Ishwaran, H., ter Brugge, P. J., Jonkers, J., Slingerland, J., & Minn, A. J. (2014). Exosome Transfer from Stromal to Breast Cancer Cells Regulates Therapy Resistance Pathways. *Cell*, 159(3), 499–513. <https://doi.org/10.1016/j.cell.2014.09.051>
- Chapuy, B., Koch, R., Radunski, U., Corsham, S., Cheong, N., Inagaki, N., Ban, N., Wenzel, D., Reinhardt, D., Zapf, A., Schweyer, S., Kosari, F., Klapper, W., Truemper, L., & Wulf, G. G. (2008). Intracellular ABC transporter A3 confers multidrug resistance in leukemia cells by lysosomal drug sequestration. *Leukemia*, 22(8), 1576–1586. <https://doi.org/10.1038/leu.2008.103>
- Chung, W.-M., Molony, R. D., & Lee, Y.-F. (2021). Non-stem bladder cancer cell-derived extracellular vesicles promote cancer stem cell survival in response to chemotherapy. *Stem Cell Research & Therapy*, 12(1), 533. <https://doi.org/10.1186/s13287-021-02600-6>
- Conigliaro, A., Costa, V., Dico, A. L., Saieva, L., Buccheri, S., Dieli, F., Manno, M., Raccosta, S., Mancone, C., Tripodi, M., Leo, G. D., & Alessandro, R. (2015). CD90+ liver cancer cells modulate endothelial cell phenotype through the release of exosomes containing H19 lncRNA. *Molecular Cancer*, 14(1), 155. <https://doi.org/10.1186/s12943-015-0426-x>
- Donnarumma, E., Fiore, D., Nappa, M., Roscigno, G., Adamo, A., Iaboni, M., Russo, V., Affinito, A., Puoti, I., Quintavalle, C., Rienzo, A., Piscuoglio, S., Thomas, R., & Condorelli, G. (2017). Cancer-associated fibroblasts release exosomal microRNAs that dictate an aggressive phenotype in breast cancer. *Oncotarget*, 8(12), 19592–19608. <https://doi.org/10.18632/oncotarget.14752>
- Doyle, L. M., & Wang, M. Z. (2019). Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. *Cells*, 8(7), 727. <https://doi.org/10.3390/cells8070727>
- Hu, J. L., Wang, W., Lan, X. L., Zeng, Z. C., Liang, Y. S., Yan, Y. R., Song, F. Y., Wang, F. F., Zhu, X. H., Liao, W. J., Liao, W. T., Ding, Y. Q., & Liang, L. (2019a). CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer. *Molecular Cancer*, 18(1), 91. <https://doi.org/10.1186/s12943-019-1019-x>
- Hu, Y.-B., Yan, C., Mu, L., Mi, Y.-L., Zhao, H., Hu, H., Li, X.-L., Tao, D.-D., Wu, Y.-Q., Gong, J.-P., & Qin, J.-C. (2019b). Exosomal Wnt-induced dedifferentiation of colorectal cancer cells contributes to chemotherapy resistance. *Oncogene*, 38(11), 1951–1965. <https://doi.org/10.1038/s41388-018-0557-9>
- Hu, Y., Yan, C., Mu, L., Huang, K., Li, X., Tao, D., Wu, Y., & Qin, J. (2015). Fibroblast-Derived Exosomes Contribute to Chemoresistance through Priming Cancer Stem Cells in Colorectal Cancer. *PLoS ONE*, 10(5), e0125625. <https://doi.org/10.1371/journal.pone.0125625>
- Huang, H., Hou, J., Liu, K., Liu, Q., Shen, L., Liu, B., Lu, Q., Zhang, N., Che, L., Li, J., Jiang, S., Wang, B., Wen, Q., Hu, L., & Gao, J. (2021). RAB27A-dependent release of exosomes by liver cancer stem cells induces Nanog expression in their differentiated progenies and confers regorafenib resistance. *Journal of Gastroenterology and Hepatology*, 36(12), 3429–3437. <https://doi.org/10.1111/jgh.15619>
- Ifergan, I., Scheffer, G. L., & Assaraf, Y. G. (2005). Novel Extracellular Vesicles Mediate an ABCG2-Dependent Anticancer Drug Sequestration and Resistance. *Cancer Research*, 65(23), 10952–10958. <https://doi.org/10.1158/0008-5472.can-05-2021>
- Jabbari, N., Karimipour, M., Khaksar, M., Akbariazar, E., Heidarzadeh, M., Mojarad, B., Aftab, H., Rahbarghazi, R., & Rezaie, J. (2020). Tumor-derived extracellular vesicles: insights into bystander effects of exosomes after irradiation. *Lasers in Medical Science*, 35(3), 531–545. <https://doi.org/10.1007/s10103-019-02880-8>
- Jella, K. K., Rani, S., O'Driscoll, L., McClean, B., Byrne, H. J., & Lyng, F. M. (2014). Exosomes Are Involved in Mediating Radiation Induced Bystander Signaling in Human Keratinocyte Cells. *Radiation Research*, 181(2), 138–145. <https://doi.org/10.1667/rr13337.1>
- Jelonek, K., Widlak, P., & Pietrowska, M. (2016). The Influence of Ionizing Radiation on Exosome Composition, Secretion and Intercellular Communication. *Protein and Peptide Letters*, 23(7), 656–663. <https://doi.org/10.2174/0929866523666160427105138>
- Kyjacova, L., Hubackova, S., Krejčíkova, K., Strauss, R., Hanzlíková, H., Dzijak, R., Imrichova, T., Simova, J., Reinis, M., Bartek, J., & Hodny, Z. (2015). Radiotherapy-induced plasticity of prostate cancer mobilizes stem-like non-adherent, Erk signaling-dependent cells. *Cell Death & Differentiation*, 22(6), 898–911. <https://doi.org/10.1038/cdd.2014.97>
- Li, I., & Nabet, B. Y. (2019). Exosomes in the tumor microenvironment as mediators of cancer therapy resistance. *Molecular Cancer*, 18(1), 32. <https://doi.org/10.1186/s12943-019-0975-5>
- Liu, L., Zhang, Z., Zhou, L., Hu, L., Yin, C., Qing, D., Huang, S., Cai, X., & Chen, Y. (2020). Cancer associated fibroblasts-derived exosomes contribute to radioresistance through promoting colorectal cancer stem cells phenotype. *Experimental Cell Research*, 391(2), 111956. <https://doi.org/10.1016/j.yexcr.2020.111956>

- Maacha, S., Bhat, A. A., Jimenez, L., Raza, A., Haris, M., Uddin, S., & Grivel, J.-C. (2019). Extracellular vesicles-mediated intercellular communication: roles in the tumor microenvironment and anti-cancer drug resistance. *Molecular Cancer*, 18(1), 55. <https://doi.org/10.1186/s12943-019-0965-7>
- Meldolesi, J. (2022). Cancer Stem Cells and Their Vesicles, Together with Other Stem and Non-Stem Cells, Govern Critical Cancer Processes: Perspectives for Medical Development. *International Journal of Molecular Sciences*, 23(2), 625. <https://doi.org/10.3390/ijms23020625>
- Mutschelknaus, L., Peters, C., Winkler, K., Yentrapalli, R., Heider, T., Atkinson, M. J., & Moertl, S. (2016). Exosomes Derived from Squamous Head and Neck Cancer Promote Cell Survival after Ionizing Radiation. *PLoS ONE*, 11(3), e0152213. <https://doi.org/10.1371/journal.pone.0152213>
- Namee, N. M., & O'Driscoll, L. (2018). Extracellular vesicles and anti-Cancer drug resistance. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1870(2), 123–136. <https://doi.org/10.1016/j.bbcan.2018.07.003>
- Nawaz, M. (2017). Extracellular vesicle-mediated transport of non-coding RNAs between stem cells and cancer cells: implications in tumor progression and therapeutic resistance. *Stem Cell Investigation*, 4(10), 83–83. <https://doi.org/10.21037/sci.2017.10.04>
- Ramakrishnan, V., Xu, B., Akers, J., Nguyen, T., Ma, J., Dhawan, S., Ning, J., Mao, Y., Hua, W., Kokkoli, E., Furnari, F., Carter, B. S., & Chen, C. C. (2020). Radiation-induced extracellular vesicle (EV) release of miR-603 promotes IGF1-mediated stem cell state in glioblastomas. *EBioMedicine*, 55, 102736. <https://doi.org/10.1016/j.ebiom.2020.102736>
- Ren, J., Ding, L., Zhang, D., Shi, G., Xu, Q., Shen, S., Wang, Y., Wang, T., & Hou, Y. (2018). Carcinoma-associated fibroblasts promote the stemness and chemoresistance of colorectal cancer by transferring exosomal lncRNA H19. *Theranostics*, 8(14), 3932–3948. <https://doi.org/10.7150/thno.25541>
- Safaei, R., Larson, B. J., Cheng, T. C., Gibson, M. A., Otani, S., Naerdemann, W., & Howell, S. B. (2005). Abnormal lysosomal trafficking and enhanced exosomal export of cisplatin in drug-resistant human ovarian carcinoma cells. *Molecular Cancer Therapeutics*, 4(10), 1595–1604. <https://doi.org/10.1158/1535-7163.mct-05-0102>
- Scholl, J. N., Dias, C. K., Muller, L., Battastini, A. M. O., & Figueiró, F. (2020). Extracellular vesicles in cancer progression: are they part of the problem or part of the solution? *Nanomedicine*, 15(26), 2625–2641. <https://doi.org/10.2217/nmm-2020-0256>
- Shen, M., Dong, C., Ruan, X., Yan, W., Cao, M., Pizzo, D., Wu, X., Yang, L., Liu, L., Ren, X., & Wang, S. E. (2019). Chemotherapy-Induced Extracellular Vesicle miRNAs Promote Breast Cancer Stemness by Targeting ONECUT2. *Cancer Research*, 79(14), 3608–3621. <https://doi.org/10.1158/0008-5472.can-18-4055>
- Su, C., Zhang, J., Yarden, Y., & Fu, L. (2021). The key roles of cancer stem cell-derived extracellular vesicles. *Signal Transduction and Targeted Therapy*, 6(1), 109. <https://doi.org/10.1038/s41392-021-00499-2>
- Sun, Z., Wang, L., Dong, L., & Wang, X. (2018). Emerging role of exosome signalling in maintaining cancer stem cell dynamic equilibrium. *Journal of Cellular and Molecular Medicine*, 22(8), 3719–3728. <https://doi.org/10.1111/jcmm.13676>
- Sun, Z., Wang, L., Zhou, Y., Dong, L., Ma, W., Lv, L., Zhang, J., & Wang, X. (2020). Glioblastoma Stem Cell-Derived Exosomes Enhance Stemness and Tumorigenicity of Glioma Cells by Transferring Notch1 Protein. *Cellular and Molecular Neurobiology*, 40(5), 767–784. <https://doi.org/10.1007/s10571-019-00771-8>
- Szatmári, T., Kis, D., Bogdándi, E. N., Benedek, A., Bright, S., Bowler, D., Persa, E., Kis, E., Balogh, A., Naszályi, L. N., Kadhim, M., Sáfrány, G., & Lumniczky, K. (2017). Extracellular Vesicles Mediate Radiation-Induced Systemic Bystander Signals in the Bone Marrow and Spleen. *Frontiers in Immunology*, 8, 347. <https://doi.org/10.3389/fimmu.2017.00347>
- Wang, L., He, J., Hu, H., Tu, L., Sun, Z., Liu, Y., & Luo, F. (2020). Lung CSC-derived exosomal miR-210-3p contributes to a pro-metastatic phenotype in lung cancer by targeting FGFR1. *Journal of Cellular and Molecular Medicine*, 24(11), 6324–6339. <https://doi.org/10.1111/jcmm.15274>
- Wang, L., Yang, G., Zhao, D., Wang, J., Bai, Y., Peng, Q., Wang, H., Fang, R., Chen, G., Wang, Z., Wang, K., Li, G., Yang, Y., Wang, Z., Guo, P., Peng, L., Hou, D., & Xu, W. (2019). CD103-positive CSC exosome promotes EMT of clear cell renal cell carcinoma: role of remote MiR-19b-3p. *Molecular Cancer*, 18(1), 86. <https://doi.org/10.1186/s12943-019-0997-z>
- Xu, J., Liao, K., & Zhou, W. (2018). Exosomes Regulate the Transformation of Cancer Cells in Cancer Stem Cell Homeostasis. *Stem Cells International*, 2018, 4837370. <https://doi.org/10.1155/2018/4837370>
- Xu, S., Wang, J., Ding, N., Hu, W., Zhang, X., Wang, B., Hua, J., Wei, W., & Zhu, Q. (2016). Exosome-mediated microRNA transfer plays a role in radiation-induced bystander effect. *RNA Biology*, 12(12), 1355–1363. <https://doi.org/10.1080/15476286.2015.1100795>
- Yang, Q., Zhao, S., Shi, Z., Cao, L., Liu, J., Pan, T., Zhou, D., & Zhang, J. (2021). Chemotherapy-elicited exosomal miR-378a-3p and miR-378d promote breast cancer stemness and chemoresistance via the activation of EZH2/STAT3 signaling. *Journal of Experimental & Clinical Cancer Research : CR*, 40(1), 120. <https://doi.org/10.1186/s13046-021-01901-1>

Zhang, Z., Xu, J., Chen, Z., Wang, H., Xue, H., Yang, C., Guo, Q., Qi, Y., Guo, X., Qian, M., Wang, S., Qiu, W., Gao, X., Zhao, R., Guo, X., & Li, G. (2020). Transfer of MicroRNA via Macrophage-Derived Extracellular Vesicles Promotes Proneural-to-Mesenchymal Transition in Glioma Stem Cells. *Cancer Immunology Research*, 8(7), 966–981. <https://doi.org/10.1158/2326-6066.cir-19-0759>

APPENDIX 3



Analysis of Bisulfite Sequencing PCR

User guide

Version 1.1.0

Marie Denoulet and Chann Lagadec

Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277 - CANTHER - Cancer Heterogeneity Plasticity and Resistance to Therapies, F-59000 Lille, France

October 7, 2022

Contents

List of Figures	3
List of Code Listings	3
1 General information	4
1.1 What ABSP does do?	4
1.2 What are the advantages of using ABSP?	4
1.3 Bisulfite Sequencing PCR	5
1.4 How does ABSP work?	7
1.5 License	9
2 How to proceed with analysis using ABSP?	12
2.1 Install ABSP	12
2.1.1 Software installations and download ABSP files	12
2.1.2 Content of the ABSP main directory	12
2.1.3 Launch the ABSP app with automated package installation	13
2.2 Open ABSP	14
2.3 Update ABSP	14
2.4 App interface	14
2.5 Individual analysis	19
2.5.1 Input files requirements	19
2.5.2 Procedure	20
2.5.3 Output report	22
2.5.4 Output files	27
2.6 Grouped analysis	29
2.6.1 Procedure	29
2.6.2 Output report	30
2.6.3 Output files	36
2.7 Multiple analyses	37
2.7.1 Input files requirements	37
2.7.2 Procedure	39
2.7.3 Output files	39
3 Complementary information	40
3.1 Some recommendations for the BSP experiment	40
3.2 Detailed workflow of ABSP individual analysis	40
3.3 Code modifications	42
3.3.1 List of reference genomes	42
3.3.2 Modify the default thresholds	42
3.3.3 Modify the plots colors and point shapes	43
4 Troubleshooting guide	46

4.1 General	46
4.2 Individual analysis	48
4.3 Grouped analysis	51
5 Acknowledgments	52

List of Figures

1	Bisulfite Sequencing PCR experimental principle	5
2	Bisulfite Sequencing PCR analysis strategies for both direct-BSP and cloning-BSP	6
3	Workflow of the ABSP analytic process	8
4	Detailed workflow of the ABSP analytic process	10
5	Diagram of the possible ways to launch ABSP analyses	11
6	RStudio interface to open ABSP app	14
7	ABSP app interface	19
8	Example of .fasta file for the reference DNA input required for ABSP individual analysis	20
9	IGV (Integrative Genomics Viewer) software window	21
10	Example of trimming plot from the output report	25
11	Diagram of output directories to locate output files from individual analysis	28
12	Visualization plots of all replicates (direct-BSP)	32
13	Visualization plots of all clones from one sample (cloning-BSP)	33
14	Visualization plots of groups (means of replicates/clones per sample)	34
15	Boxplots and methylation profile plots	36
16	Diagram of output directories to locate output files from individual analysis	36
17	Detailed workflow of the individual analysis	41

List of Code Listings

1	List of reference genomes displayed in the drop-down lists	42
2	Genome name selected by default in the individual analysis tab.	42
3	Genome name selected by default in the grouped analysis tab.	43
4	Default thresholds in the individual analysis script	43
5	Default thresholds in the grouped analysis script	44
6	Colors and shapes setting for plots in the grouped analysis script	45

1 General information

1.1 What ABSP does do?

ABSP, standing for "Analysis of Bisulfite Sequencing PCR", is an R-based tool to analyze CpG methylation profiles using data from Bisulfite Sequencing PCR (BSP) experiment results. It was developed to help researchers to estimate and compare methylation percentages of a DNA region studied using BSP experiments. It provides a complete automated workflow, from trace file sequencing results to data visualization and statistics.

1.2 What are the advantages of using ABSP?

- **A complete workflow.** ABSP uses as input the chromatogram trace files as the sequencing results, and through a two-steps analysis, it (1) computes the methylation percentages of individual samples after validating the sequencing quality and (2) gathers the methylation levels from all samples to summarize methylation data, generate publication-ready figures and perform comparative statistics to answer to the experiment hypothesis on the DNA methylation differences between conditions.
- **A fully automated process.** ABSP uses a shiny app on R to provide a user-friendly interface. To launch the analytic process the user is guided to provide the required inputs and can launch the desired analysis with one click. For each analysis, an HTML report file is generated to visualize the results and keep a record of them. Additionally, output files, such as tables and figures, are automatically saved in the corresponding result folders. For an even more automated use, several analyses can be launched with the help of pre-filled input tables (spreadsheet document to fill) in the special tab "Multiple analyses", which is useful for large amounts of samples.
- **Analyses of both direct-BSP and cloning-BSP sequencing data.** ABSP can analyze results from both BSP methods. No existing tool is currently able to analyze both. It allows continuity in the experiment analytic process, as the direct-BSP approach can be performed before cloning the PCR products to have preliminary insights on DNA methylation, and then further confirmed/validated using cloning-BSP.
- **Accessible and flexible.** ABSP is coded using R, a crossed-platform tool language increasingly used in biology research, making it very accessible to any researchers. Additionally, for researchers accustomed to R coding, as the entire scripts are provided, ABSP is fully upgradeable. Also, we provide specific guidelines to easily modify some features to adapt ABSP to experiment needs, such as adjusting quality thresholds or changing graphical parameters (see section 3.3 Code modifications at page 42).

1.3 Bisulfite Sequencing PCR

The Bisulfite Sequencing PCR (BSP) is an experimental technique aiming to estimate methylation levels of CpG sites on a specific DNA region of interest, among a population of DNA molecules. The method was originally developed by Frommer *et al.* in 1992¹ and Clark *et al.* in 1994² and was named BSP in opposition to the methylation-specific PCR (MSP) method by Li *et al.* in 2002.³

This method is composed of three steps, described in figure 1:

1. A DNA bisulfite conversion
2. A PCR amplification and an optional cloning
3. A sequencing of either PCR products or individual subclones

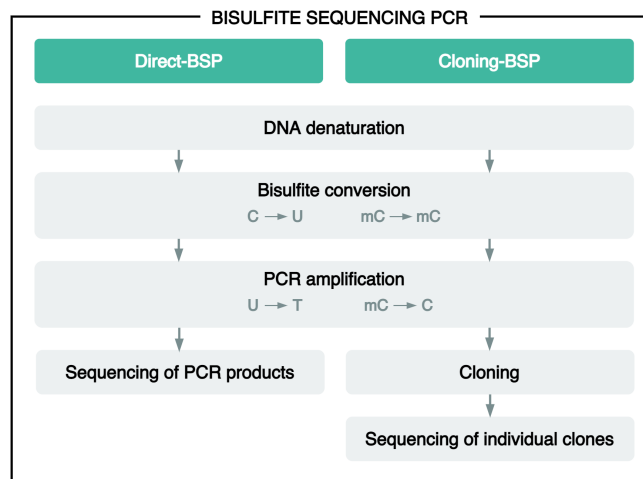


Figure 1. Bisulfite Sequencing PCR experimental principle.

Two approaches of BSP could be used (Figure 1). The *direct-BSP* method is characterized by the direct sequencing of PCR products, whereas the *cloning-BSP* consists of cloning PCR products within a specific vector and sequencing several individual clones.

¹M Frommer *et al.* "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands." In: *Proceedings of the National Academy of Sciences* 89.5 (1992), pp. 1827–1831. ISSN: 0027-8424. DOI: [10.1073/pnas.89.5.1827](https://doi.org/10.1073/pnas.89.5.1827).

²S J Clark *et al.* "High sensitivity mapping of methylated cytosines." In: *Nucleic acids research* 22.15 (1994), pp. 2990–7. ISSN: 0305-1048. DOI: [10.1093/nar/22.15.2990](https://doi.org/10.1093/nar/22.15.2990).

³Long-Cheng Li and Rajvir Dahiya. "MethPrimer: designing primers for methylation PCRs". In: *Bioinformatics* 18.11 (2002), pp. 1427–1431. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/18.11.1427](https://doi.org/10.1093/bioinformatics/18.11.1427).

As described in Figure 2, the analysis strategies of these two sub-methods are different on some points.

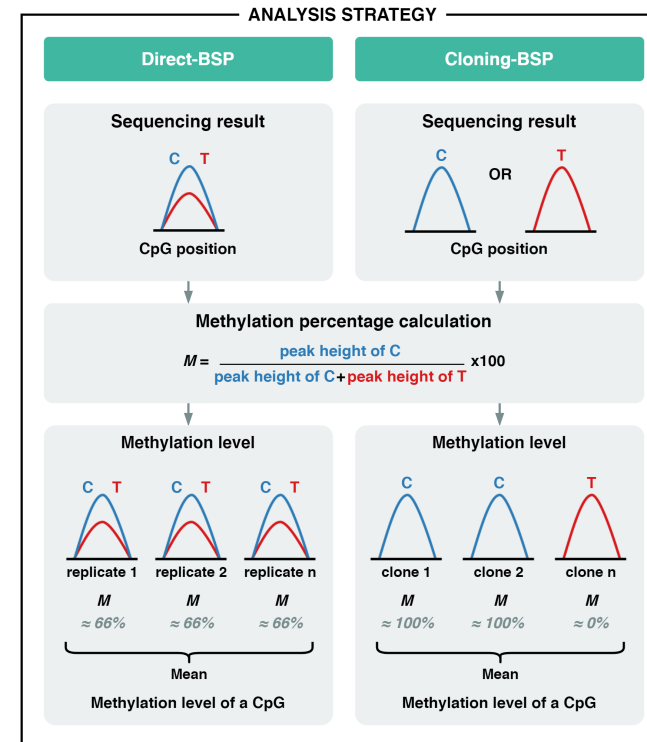


Figure 2. Bisulfite Sequencing PCR analysis strategies for both direct-BSP and cloning-BSP.

First, in **direct-BSP**, a mix of DNA molecules with different unknown methylation statuses are sequenced, thereby, at each CpG site, two base signals can co-exist: the methylated signal (C) and the unmethylated signal (T). By calculating the signal ratio, the methylation level of a CpG in the DNA population can directly be estimated, but for reproducibility and statistical significance purposes, it still needs to be repeated in several biological replicates to obtain the final methylation level of a CpG (direct-BSP results are considered less quantitative than cloning-BSP ones).

Secondly, in **cloning-BSP**, as each clone represents only one PCR product, the methylation status of a CpG can be either methylated (C) or unmethylated (T), thereby only one of these two signals can exist. Accordingly, the signal ratio can

only give either 0% or 100% of methylation (partial methylation is considered biased results), revealing the methylation status. For each CpG, the proportion of clones with a methylated status reveals the methylation level of the CpG in the original DNA population, in general, a minimum of 10 clones is recommended to have a 10% accuracy of the methylation level.

1.4 How does ABSP work?

As a first step, each BSP sequencing result is defined by a combination of experiment information (figure 3):

- **Sequence.** The sequence identifier refers to a unique amplicon sequence produced by the BSP experiment, using a unique set of primers. For example, if several regions of a gene are analyzed by BSP, as each region corresponds to a unique amplicon they must have distinct sequence names (e.g. *CDH1promoter* and *CDH1exon1*; *CDH1-1*, *CDH1-2*, and *CDH1-3*). Make sure the sequence name is strictly identical for all samples of the same sequence.
- **Collection.** The collection corresponds to a separation of samples above groups. Samples from different collections can not be compared, even if they belong to the same group. For example, collections can be different cell lines, organs, or patients, in which the same groups are compared but not between the different collections. To compare these types of samples, consider them as groups. Make sure the collection name is strictly identical for all samples of the same collection.
- **Group.** The group corresponds to the condition that will be compared with other groups/conditions in the grouped analysis. For example, groups can be the "control" and "treated" conditions. Make sure the group name is strictly identical for all samples of the same group.
- **Replicate.** Information has to be provided only when using the direct-BSP approach. The replicate number refers to the repetition identifier number of the sequencing. To have robust and reproducible data and to perform comparative statistics, each sample needs to be sequenced at least three times (in both directions).
- **Clones.** Information has to be provided only when using the cloning-BSP approach. To estimate the methylation levels among the DNA population, the methylation statuses of several individual clones needs to be sequenced. The ratio of methylated and unmethylated clones for a CpG position will give the methylation level estimation.
- **Sequencing files.** The sequencing read of each sample is performed in both directions, using a forward primer and a reverse primer.
- **Experiment.** The term "experiment (data)" refers to the unique combination of collection, group, and replicate or clone, for a specific sequence, corre-

sponding to the sample information for the sequencing read.

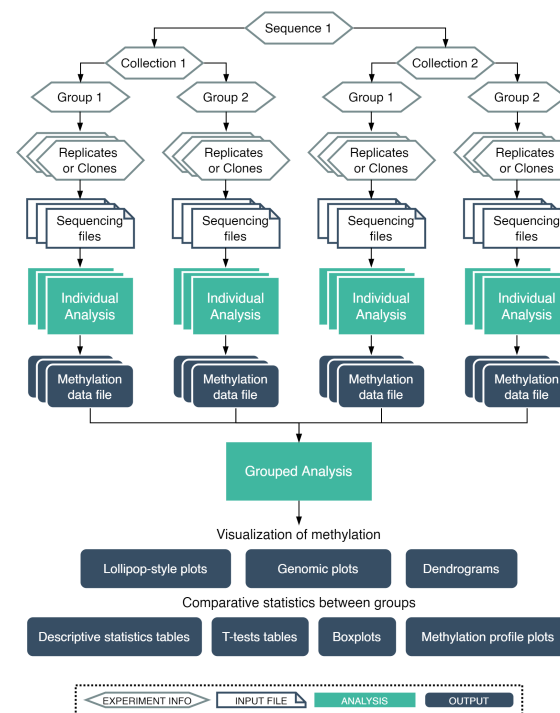


Figure 3. Workflow of the ABSP analytic process.

Then, the ABSP analysis is divided into two steps, corresponding to two scripts (R markdown scripts), as illustrated in the workflow in figure 3.

- **Individual analysis.** For each individual experiment point, two sequencing .ab1 files (one from forward direction, one from the reverse direction) are used as input for the individual analysis. First, the sequencing reads are trimmed based on quality to get the correct sequence for alignments with the reference DNA input. Then, results from the alignments go through a quality control step to check for mismatches, gaps, length of aligned sequences, and bisulfite conversion rates (calculated on cytosines outside CpG that should be thymines). If the results are defined as correct, the methylation levels of CpG can be calculated and visualized on a genomic plot. Several output files (e.g. chromatograms, sequences, tables) are saved in folders, especially the methylation data file as a result of the individual sequencing experiment analyzed.

- **Grouped analysis.** All methylation data files from the same sequence are gathered by the grouped analysis. First, a preprocessing step is performed to organize data. Then, visualization plots, lollipop-style plots, and genomic plots (with associated clustering dendrograms) are generated to view methylation data differences. Finally, a statistical analysis is performed, descriptive statistics tables and Student's t-test p-values tables are generated, as well as boxplots with t-test p-values and methylation profile plots with Kruskal–Wallis test p-values, to display the significant methylation differences.

For more details, the inputs, processes, and outputs of these two steps, individual analysis, and grouped analysis, are displayed in the [figure 4](#), where the 3 tabs of ABSP are represented: "Individual analysis", "Grouped analysis", and "Multiple analyses".

The two steps, individual analysis and grouped analysis, require the manual entry of input data in the corresponding tab, and therefore only one analysis can be launched at the same time ([figure 5](#)). An additional tab has been implemented to launch multiple analyses all at once. This multiple analyses tab can be used to launch either several individual analyses and/or several grouped analyses at the same time, by using tables (.xlsx files or .csv files) as input instead of the manual entry of input for unique analysis.

1.5 License

ABSP, Analysis of Bisulfite Sequencing PCR

Copyright © 2022 by the CANTHER laboratory, France (absp@univ-lille.fr)

Released under the GPL-3 license.

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <https://www.gnu.org/licenses/>.

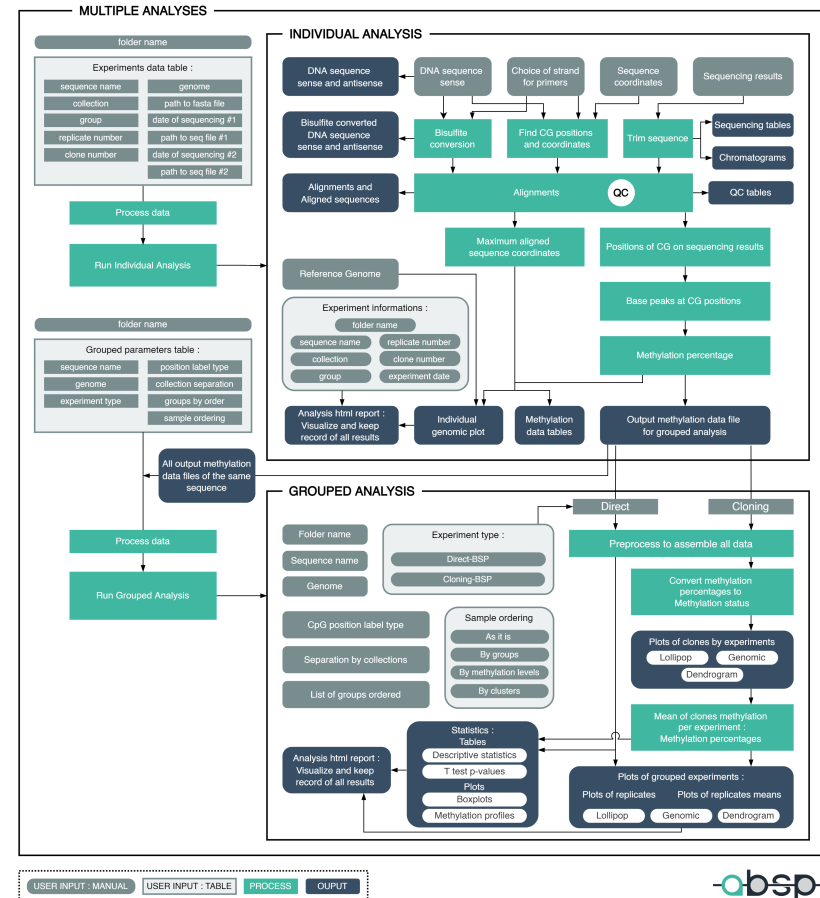


Figure 4. Detailed workflow of the ABSP analytic process.

2 How to proceed with analysis using ABSP?

2.1 Install ABSP

2.1.1 Software installations and download ABSP files

- Install R: <https://www.r-project.org/>
- Install RStudio: <https://www.rstudio.com/>
- Download the ABSP zip folder on github: <https://github.com/ABSP-methylation-tool/ABSP> and unzip the folder to access files

2.1.2 Content of the ABSP main directory

The main ABSP folder is organized as follows:

- **documents** folder for documents available to the user
 - List of BSgenomes.xlsx file listing the available genomes
 - multiple_grouped_parameters_table.xlsx table of inputs to launch multiple grouped analyses
 - multiple_grouped_parameters_table.ods table of inputs to launch multiple grouped analyses
 - multiple_individual_analyses_table.xlsx table of inputs to launch multiple individual analyses
 - multiple_individual_analyses_table.ods table of inputs to launch multiple individual analyses
- **reports** folder for analysis reports
- **results** folder for analysis results (data, tables, graphics...)
- **scripts** folder for scripts and associated files required to run analysis
 - ABSP_functions.R R script providing the necessary functions for ABSP
 - ABSP_grouped_analysis.RMD R markdown script of grouped analysis
 - ABSP_individual_analysis.RMD R markdown script of individual analysis
 - custom.css CSS script for custom theme settings of the .html report files
 - logo.svg ABSP logo vector image

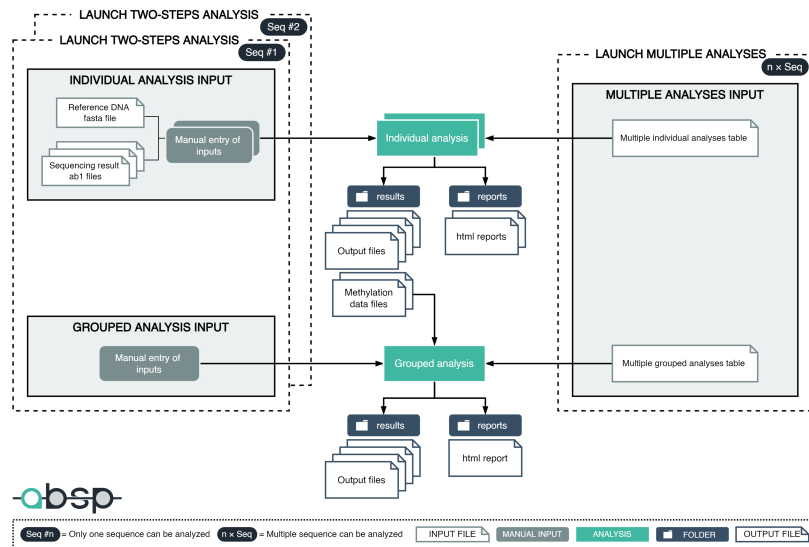


Figure 5. Diagram of the possible ways to launch ABSP analyses.

- **WWW** folder for files necessary for the shiny app
 - [ABSP - Analysis.svg](#) diagram of ABSP analysis strategy
 - [ABSP - BSP.svg](#) diagram of BSP experiment principle
 - [ABSP - fasta file.png](#) image of a reference sequence .fasta file example
 - [ABSP - Launch analysis.svg](#) diagram of the different ways to launch ABSP analyses
 - [ABSP - Workflow simple.svg](#) diagram of the BSP workflow
 - [custom_app.css](#) CSS script for custom theme settings of the app interface
 - [logo.svg](#) ABSP logo vector image
- [ABSP RProject.Rproj](#) R project file
- [ABSP User Guide.pdf](#) reference manual to use ABSP
- [app.R](#) shiny app file

For ABSP to function properly, all the aforementioned files must be downloaded and present in the ABSP main folder with the same structure.

Make sure not to rename, move or delete the provided folders and files. If you want to reorganize files or folders it is better to copy to other directories than to modify the files. However, new folders can be added to the ABSP main folder without causing issues.

2.1.3 Launch the ABSP app with automated package installation

- Open the [ABSP Rproject.Rproj](#) file with RStudio.
- Open the [app.R](#) file with RStudio.
- *Optional:* Find the "Run App" button in the upper right corner, click on the arrow right next to it and select "Run external", to open the app with the default web browser instead of a RStudio window.
- Click on the "Run App" button to launch the app (Figure 6).
- A pop-up window should appear if the `shiny` package was not already installed on your device, click on "Yes" to accept the `shiny` installation.
- Once this procedure is done, the package installation should start and it might take a few minutes. Please note that installation on a Linux device can be quite long (approx. 40 min) compared to Windows/MacOS devices.
- If a message in the console asks to update packages, respond positively to update all packages.
- Once the app is opened, the packages required for ABSP should be completely installed, therefore analyses can be carried out from the app interface.

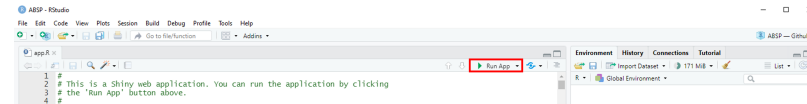


Figure 6. RStudio interface to open ABSP app.

2.2 Open ABSP

- Open the [ABSP Rproject.Rproj](#) file with RStudio.
- Open the [app.R](#) file with RStudio.
- Click on the "Run App" button to launch the app.

To launch the different analyses refer to the following sections below which describes the individual analysis, grouped analysis, and multiples analyses (Figure 5).

2.3 Update ABSP

Download the latest version of ABSP files on GitHub: <https://github.com/ABSP-methylation-tool/ABSP>.

Use the new ABSP main directory (latest version) to open ABSP.

You can use your previous input files, but make sure they are compatible with the new version (e.g. *the multiple_individual_analyses_table* and *multiple_grouped_analyses_files* templates for multiple launches of analyses might have been changed).

As the R project and app are from the new ABSP main directory (latest version), the results and reports will be saved in this directory.

2.4 App interface

The implemented shiny app provides a user interface, opening in a web browser, aiming to facilitate the input entry and the launching of analyses for users.

The interface contains 4 tabs, displayed in Figure 7:

1. A **main** tab.
2. An **individual analysis** tab.
3. A **grouped analysis** tab.
4. A **multiple analyses** tab.

The screenshot shows the obbsp web application interface. At the top, there are navigation tabs: 'Main', 'Individual analysis', 'Compound analysis', and 'Multiple analyses'. The 'Sample information' tab is selected, showing a form with the following fields: 'Sample name' (Example data), 'Reference DNA sequence' (CH11), 'Enter collection name' (CH11), 'Enter group name' (High-methylated DNA), 'Reference DNA sequence' (CH11 ABSP, Enter clone number: 1), and 'Sequencing results' (Select date of sequencing #1: 2022-05-02, Select date of sequencing #2: 2022-05-02). Below the form, there are instructions for each field, such as 'Reference DNA sequence' and 'Sequencing results'. The 'Sequencing results' tab is also visible, showing 'Select date of sequencing #1' and '#2' with dropdown menus for 'CH11' and 'CH12'.

16

Figure 7 Panel 2. ABSP app individual analysis tab.

The screenshot shows the 'What is ABSP?' section of the obbsp web application. It includes a 'Please, cite' section with a citation for Chavakis et al. (2022). Below this, there is a 'Resources' section with links to the ABSP GitHub repository and a preprint. At the bottom, there is a 'Contact' section with an email address.

15

The diagram illustrates the obbsp workflow. It starts with 'Bisulfite treatment' (CH11, CH12, CH13, CH14, CH15, CH16, CH17, CH18, CH19, CH20, CH21, CH22, CH23, CH24, CH25, CH26, CH27, CH28, CH29, CH30, CH31, CH32, CH33, CH34, CH35, CH36, CH37, CH38, CH39, CH40, CH41, CH42, CH43, CH44, CH45, CH46, CH47, CH48, CH49, CH50, CH51, CH52, CH53, CH54, CH55, CH56, CH57, CH58, CH59, CH60, CH61, CH62, CH63, CH64, CH65, CH66, CH67, CH68, CH69, CH70, CH71, CH72, CH73, CH74, CH75, CH76, CH77, CH78, CH79, CH80, CH81, CH82, CH83, CH84, CH85, CH86, CH87, CH88, CH89, CH90, CH91, CH92, CH93, CH94, CH95, CH96, CH97, CH98, CH99, CH100). This is followed by 'Sequencing' (CH11, CH12, CH13, CH14, CH15, CH16, CH17, CH18, CH19, CH20, CH21, CH22, CH23, CH24, CH25, CH26, CH27, CH28, CH29, CH30, CH31, CH32, CH33, CH34, CH35, CH36, CH37, CH38, CH39, CH40, CH41, CH42, CH43, CH44, CH45, CH46, CH47, CH48, CH49, CH50, CH51, CH52, CH53, CH54, CH55, CH56, CH57, CH58, CH59, CH60, CH61, CH62, CH63, CH64, CH65, CH66, CH67, CH68, CH69, CH70, CH71, CH72, CH73, CH74, CH75, CH76, CH77, CH78, CH79, CH80, CH81, CH82, CH83, CH84, CH85, CH86, CH87, CH88, CH89, CH90, CH91, CH92, CH93, CH94, CH95, CH96, CH97, CH98, CH99, CH100). The 'Analysis' process is divided into 'Bisulfite conversion rate' and 'Methylation level'. The 'Bisulfite conversion rate' is calculated as $\frac{C}{C+T}$ and the 'Methylation level' is calculated as $\frac{C}{C+T}$. The diagram also includes a 'Please, cite' section with a citation for Chavakis et al. (2022) and a 'Resources' section with links to the ABSP GitHub repository and a preprint.

Figure 7 Panel 1. ABSP app main tab.

obsp Analysis of Bisulfite Sequencing PCR

Main Individual analysis **Grouped analysis** Multiple analysis

Select existing folder
Example data

Select individual analysis table file
Example data
multiple_individual_analysis_table_CDH13A

Select grouped analysis table file
Example data
multiple_grouped_analysis_table_CDH13A

Upload file: multiple_individual_analysis_table_CDH13A
multiple_grouped_analysis_table_CDH13A

Run analyses

Principle Diagram of analysis launch

Automated launch of multiple analyses
Several analyses can be launched from this tab. In only one click, using pre-defined tables containing the input results.

How to proceed ?

1. Fill over the table input tables below with the sample information and the choice of parameters (you can open the table files with the button below).
2. In the first panel, select an existing folder where the ABSP results folder to be used as all the analysis results.
3. In the second panel, select the 'Create new folder' entry and enter the name of the new folder in the text input. Note that the link and icon will appear in the report file name.
3. In the left panel, select your direct table as input.

Note: The sample data table can be multiple individual analysis for the parameter's table for multiple grouped analyses or be provided at the same time to launch the individual analysis followed by the grouped analysis, or a table of sample information can be used to launch grouped analyses, even individual analysis or grouped analyses.

After the input analyses will be launched, the input file will be opened in the report file.

Note: The table files located in the document folder. They are files named as either 'job' or 'job'.

Sample data for individual analyses
The table requires the individual analysis inputs.

Open multiple individual analysis table file

Parameters for the grouped analysis
The table requires the proper analysis types.

Open multiple grouped analysis table file

18

obsp Analysis of Bisulfite Sequencing PCR

Main Individual analysis **Grouped analysis** Multiple analysis

Experiment information **Run parameters**

Select folder
Example data

Select sequence
CHH1

Select genome
Reference: hg38/UCSC/hg19

Select experiment type
Direct BSIP Chasing BSIP

Select parameters
Select position labels for peaks
 CpG islands CpG islands' None
Collection operation
 Separate peak by collection
 Indicate the order of groups (all must be selected)
Highly methylated DNA (Low methylated DNA)
Select the type of sample ending for peak (multiple choice allowed)
All the Bisopur By methylation levels Bisulites

Run grouped analysis

Experiment information **Run parameters**

Select folder
Select an existing folder within the ABSP results folder to reference the previously generated methylation data files after individual analyses, and to locate the newly generated results of the grouped analysis.

Select sequence
Select an existing folder corresponding to the name of your sequence, to reference the previously generated methylation data files after individual analyses, and to locate the newly generated results of the grouped analysis.

Select genome
Select the reference genome. Only used to display the genome sequence in the genome plots.

Select experiment type
Select the experiment type: either 'Direct BSIP' or 'Chasing BSIP' for input or retrieval of methylation data and analysis specification.

17

Figure 7 Panel 3. ABSP app grouped analysis tab.

Figure 7. ABSP app interface. **Panel 1.** Main tab or home page of the app, it provides key links and resources for the users. **Panel 2.** Individual analysis tab, it serves to launch the analysis of a single sample using two sequencing files, one for each direction, to obtain CpG methylation percentages. On the left, input information and files must be provided in the dedicated text and file upload boxes. **Panel 3.** Grouped analysis tab, it serves to launch the analysis of all sample methylation results after their individual analysis to compare CpG methylation percentages between groups. On the left, input information must be selected in the selection boxes. **Panel 4.** Multiple analyses tab, this tab can be used to launch multiple individual and/or grouped analyses at once, using pre-filled input tables that must be provided in the upload file boxes.

The **main** tab exposes general information and resources. The **individual analysis** and **grouped analysis** tabs are dedicated to the launching of the individual analysis and grouped analysis, respectively. They are divided into two parts, one left panel with input boxes and the "Run analysis" buttons, and one right panel containing subtabs that provide all the necessary guidelines to fill the input boxes. The **multiple analyses** tab aims to launch both individual analyses and grouped analyses through the upload of a pre-filled table file containing all the analyses inputs. It is also composed of a left input panel and a right panel with input guidelines. This tab procures a further automated way to run ABSP analyses, more details are given in section 2.7 Multiple analyses at page 37.

2.5 Individual analysis

2.5.1 Input files requirements

Sequencing result .ab1 files As input, ABSP requires the chromatogram trace file (.ab1) from the sequencing run (Sanger) using the bisulfite converted DNA PCR products as templates. It is highly recommended to have both directions sequenced, by a primer on each side of the PCR product: a forward primer and a reverse primer. Although, the analysis can be run using only one trace file instead of both. The directions do not need to be specified as the analysis will determine it automatically.

Reference DNA sequence and information .fasta file ABSP also requires a .fasta file containing information about the reference DNA. A .fasta file is composed of a **header** and a **body**. The header must contain both the **genomic coordinates** of this sequence and the **choice of strand**, the one used for primer design, the one that will be amplified by PCR. Indeed, as both strands are no longer complementary after bisulfite conversion, the primers have to be designed on only one bisulfite converted strand as template DNA. The body must contain the **nucleotide sequence** of the reference DNA from the **plus strand** (upper/sense strand) of the genome.

Formats for the .fasta file header:

- The **genomic coordinates** must be written in the format `chr#.#####-#####`

(e.g. `chr16:68771087-68771462`).

- The **choice of strand amplified** must be either "`primers=plus`" or "`primers=minus`". If none of these character strings are present in the .fasta file, by default the plus strand is chosen.

An example of a .fasta file content is depicted in Figure 8. Note that any other information in the header, such as the sequence name, for example, can be added without consequences if they do not interfere with the previously described formats.

```
> CDH1 chr16:68770900-68771299 primers=plus
CCAAGTGTAAAAGCCCTTCTGATCCCAGGTCTTAGTGAGCCACCGCGGGGCTGGGATTGGAACCCAGTGGAAATCAGAAC
CGTGCAGGTCCCATAAACCCACCTAGACCCTAGCAACTCCAGGCTAGAGGGTACCCTGCTATGCGAGGCCGGTGGGCGG
GCCGTCAGTCCGCCCTGGGGAGGGTCCGGCTGCTGATTGGCTGTGGCCGGCAGGTGAACCCTCAGCAATCAGCGGTA
CGGGGGCCGTGCCTCCGGGGCTCACCTGGCTGCAGCCACGCACCCCTCTCAGTGGCGTCGGAACGCAAGCACCTGTG
AGCTTGCAGGAGTCACTCAGACTCCAGCCCGCTCCAGCCCGCCGACCCGACCCCGCGCTGCCCTCGC
```

Figure 8. Example of .fasta file for the reference DNA input required for ABSP individual analysis.

As a recommendation, to help with the creation of the .fasta file, the **IGV (Integrative Genomics Viewer)**⁴ software can be easily used to navigate the genome and to get the nucleotide sequence of a specific region, alongside with its genomic coordinates. After navigating on the genome, the current viewed region can be added as a region of interest by selecting the "Regions" > "Region Navigator" > "Add" (Figure 9). In this "Regions of Interest" panel, the added region will appear in the list of regions, its coordinates can be adjusted and it can be annotated with a description. By right-clicking on a region from this list, a context menu appears and two options can be selected, "Copy Sequence" or "Copy Details". The first one copies to the clipboard the nucleotide sequence of the region, and the second one copies the genomic coordinates (in the correct format for the .fasta file) as well as the description associated with the region. Make sure to properly verify the first and last nucleotides, as there can be a one nucleotide difference between the coordinates and the actual sequence.

2.5.2 Procedure

In the **individual analysis** tab, the left panel is to launch analysis, and the right panel provides entry information.

Experiment information

- **Select an existing folder or enter a new folder name.** Located in the `AB-SP/results` folder, all of the analysis results will be saved in this folder. Having different folders of results can be used to separate the different analyses by projects, experiments, or users. Note that the six first letters of the folder name will appear in the report file name.

⁴<https://software.broadinstitute.org/software/igv/>

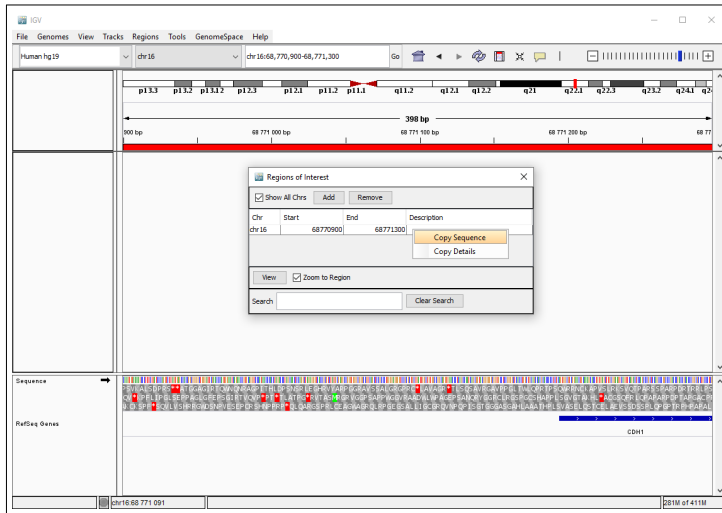


Figure 9. IGV (Integrative Genomics Viewer) software window.

- **Select an existing sequence folder or enter a new sequence name.** The sequence folder is located in the previously selected folder. All of the analysis results will be saved in this folder corresponding to the same sequence. Note that this sequence name will be used in output files (tables, plots...) to refer to the sequence.
- **Enter the collection name.** The collection corresponds to a separation of samples above groups (*details in section 1.4 How does ABSP work? at page 7*). Make sure the collection name is strictly identical for all samples of the same collection.
- **Enter the group name.** The group corresponds to the condition to compare (*details in section 1.4 How does ABSP work? at page 7*). Make sure the group name is strictly identical for all samples of the same group.
- **For direct-BSP only: Enter replicate number.** In the case of direct sequencing of PCR products only, the replicate number corresponds to the repetition identifier of the sample (*details in section 1.4 How does ABSP work? at page 7*).
- **For cloning-BSP only: Enter clone number.** In the case of clone sequencing only, the clone number corresponds to the identifier number of each clone from the same condition (*details in section 1.4 How does ABSP work? at page 7*).

Reference DNA sequence

- **Select the reference genome.** It will only be used to display the genomic sequence in the genomic plot. Make sure to click on the "Pre-install genome" button if the selected genome is used for the first time. Only a short list of genomes is displayed in the drop-down list but more genome assemblies are available. Go to section 3.3 *Code modifications at page 42* to get information on how to add another genome in the drop-down list. The complete list of available genomes can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.
- **Select .fasta file of reference DNA sequence.** As described above in the Input files requirements section, the .fasta file of the reference DNA sequence needs to be selected from your folders.

2.5.3 Output report

The HTML report file of the analysis is automatically saved in the *reports* folder in the ABSP directory.

Header First, in the top panel, the information about the sample experimental conditions is displayed in a table.

- Folder name
- Sequence name
- Collection
- Group
- Replicate number (for direct-BSP only)
- Cloning number (for cloning-BSP only)
- Date of sequencing #1
- Date of sequencing #2
- Date of analysis
- Prefix of output files

Reference DNA This tab summarizes all the data computed from the reference DNA sequence .fasta file.

- **Reference DNA sequence** General information on the sequence (name, strand used for primer design, length, and genomic coordinates) and genomic sequence from plus strand (given by the reference DNA .fasta file) and minus strand (reverse complement).
- **Localization of CG dinucleotides** Detection of CpG sites on the reference DNA (on plus and minus strand) with attribution of the CG number, from 1 to n on the plus strand.
- **Bisulfite converted sequences** Sequences of reference DNA after theoretical bisulfite conversion (CpG sites considered as methylated). The bisulfite

conversion is performed on the strand used for PCR primer design, as only this strand is amplified during PCR. The PCR regenerates the opposite strand, corresponding to the reverse complement of the bisulfite converted DNA template.

Sequencing trimming This tab summarizes the trimming of sequencing reads based on quality. Two parameters are used: the Phred quality score of each base retrieved from the sequencing file, and the mixed base peak ratio.

- **Summary** In the first tab, the default thresholds used to trim the sequencing results are displayed.
 - Minimum length of the trimmed sequence (default is 30 bp)
 - Minimum Phred quality score (default is 30, corresponding to a base-calling error probability of 0.001%)
 - Minimum ratio of primary peak (default is 0.75)
 - Minimum percentage of non-mixed positions (default is 75%)

Below the threshold table, the trimming summary for both sequencing reads is displayed and indicates whether or not the trimming was successful (correct trimmed sequence quality) or failed (incorrect trimmed sequence quality).

- **Details per sequencing**
 - **Raw sequence** The sequence, chromatogram, and data table of the sequencing results are displayed.
 - **Quality report** The first trimming is based on the base-calling quality as it uses the Phred quality scores of each base to find the best sequence to trim. This step is provided by the [SangeranalyseR package](#)^{5,6}. The thresholds and the results of this quality trimming are displayed.
 - **Mixed base peak report** The second trimming is based on the primary peak ratio over the other peaks for each position. At each position, the signal ratio of the primary peak is computed using the peak height values for each base, with formula: if $peak_C > \{peak_A, peak_T, peak_G\}$

$$Primary\ peak\ ratio = \frac{peak_C}{peak_A + peak_T + peak_G + peak_C}$$

If the ratio is above the threshold (default is 0.75), the position is considered non-mixed; if the ratio is below the threshold, the position is considered mixed. All the possible trimmed sequences are obtained by selecting the sequence between n (from 3 to 15) consecutive non-mixed positions. For each one of the possible trimmed sequences, the percentage of non-mixed positions is calculated. Among those, the trimmed sequence which is selected corresponds to the one with a

⁵<https://sangeranalyser.readthedocs.io/en/latest/index.html>

⁶Kuan-Hao Chao et al. "sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R". in: *bioRxiv* (2020), p. 2020.05.18.102459. doi: [10.1101/2020.05.18.102459](https://doi.org/10.1101/2020.05.18.102459).

percentage of non-mixed positions above the threshold (default is 75%) with the minimum of consecutive non-mixed positions at extremities (this number is displayed).

- **Trimming plot** The two previous report steps give two different trimmed sequences that can be viewed on the trimming plot (Figure 10). The top panel represents a dot plot of the Phred quality score per position, values in green are above the threshold and values in red below, in which the start and end positions of the trimmed sequence are represented by orange vertical lines. The second panel is also a dot plot but it represents the primary peak ratio per position, values in green are above the threshold (considered as non-mixed) and values in red below (considered as mixed), in which the start and end positions of the trimmed sequence are represented by cyan vertical lines. In the last panel, the two trimmed sequences are represented in the same color, orange, and cyan. The raw sequence is displayed in red. The overlapping of the two previous trimmed sequences gives the final trimmed sequence, in green, which is the one kept for the rest of the analysis and corresponds to the information given in the summary tab and the final trimmed sequence tab.
- **Final trimmed sequence** The sequence, chromatogram, and data table of the final trimmed sequencing results are displayed.

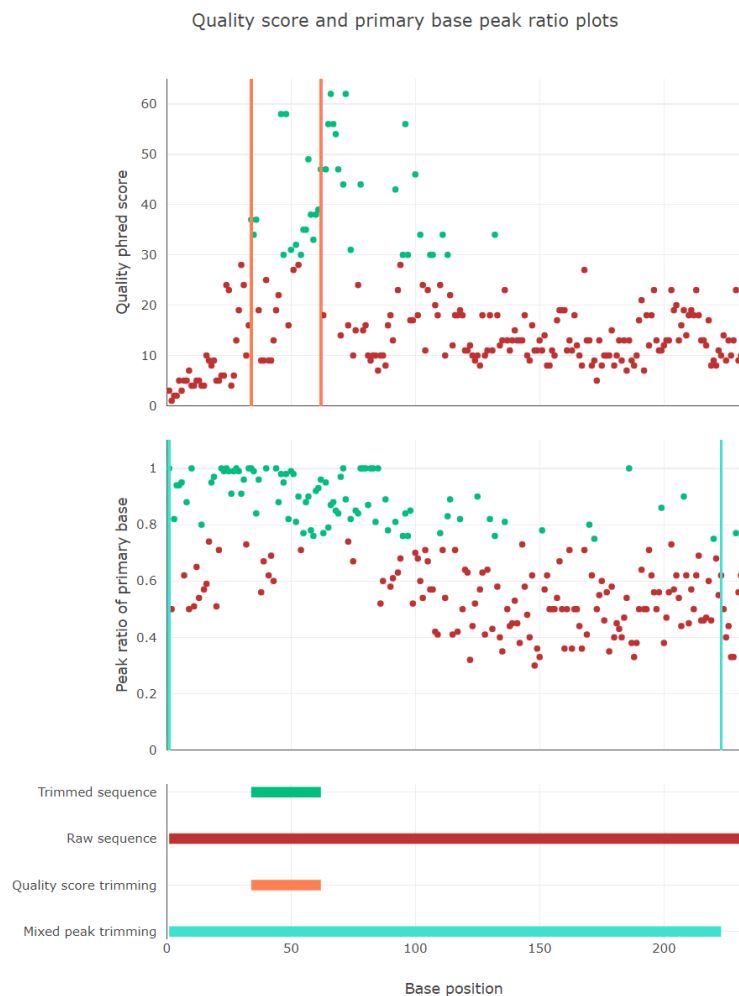


Figure 10. Example of trimming plot from the output report.

Alignments In the first two tabs, each sequencing read is aligned with either the sense sequence (bisulfite converted sequence from the template strand) as if it is a forward sequencing, or the antisense sequence (reverse complement of bisulfite converted sequence from the template strand) as if it is a reverse sequencing. The direction of each sequencing result is determined based on the aligned sequence length: the alignment which gives the longest aligned sequence is considered the correct one. If for one sequencing read the aligned sequences are equal between alignment as forward and as reverse, the direction determination depends on the other one. The two last tabs display the correct alignments and aligned sequences.

Quality control This tab summarizes the quality of the trimmed sequencing result aligned with the reference DNA.

- **Summary** In the first tab, the default thresholds are used to control the quality of the aligned sequencing result.
 - Minimum length of the aligned sequence (default is 30 bp)
 - Minimum identity percentage of alignment (default is 75%)
 - Minimum of bisulfite conversion rate mean (default is 0.90, corresponding to 90% conversion efficacy)

Below the threshold table, the quality summaries for both sequencing results are displayed and indicate whether or not the aligned sequencing results have a sufficient quality (correct) or insufficient (incorrect) relative to thresholds.

- **Mismatch positions** For both sequencing reads a table indicates the mismatched positions and nucleotides, on both the sequencing read and reference DNA sequence.
- **Insertions/deletions** For both sequencing reads a table indicates the insertions/deletions (gaps) found in either the sequencing read or the reference DNA sequence.
- **Conversion rates** For both sequencing reads a table indicates the bisulfite conversion rate for each cytosine outside a CpG in the aligned sequence. The first column corresponds to the identifier number of the cytosine on the reference DNA sequence, in the second one is displayed its position on the reference DNA sequence, in the third one its position on the trimmed sequencing result, and in the fourth its position on the raw sequencing result. The position matching is obtained thanks to the alignment of sequences. For each position on the raw sequencing result, the peak height values of signals are extracted and used to compute the conversion rate, with the following formulas (for forward and reverse sequencing respectively):

$$\text{Bisulfite conversion rate} = \frac{\text{peak}_T}{\text{peak}_C + \text{peak}_T}$$

$$\text{Bisulfite conversion rate} = \frac{\text{peak}_A}{\text{peak}_G + \text{peak}_A}$$

For each position a bisulfite conversion is obtained. The mean of rates from

all positions is indicated in the summary table in the first tab, as well as standard deviation.

- **Maximum aligned sequence** The maximum aligned sequence corresponds to the sequence covered by at least one of the two sequencing results. Its information such as its length, its coordinates, and its nucleotide sequence are displayed.

Methylation

- For both sequencing reads, the tables of computed **methylation percentages** are displayed. The first column corresponds to the CpG site identifier number on the reference DNA sequence (list of all CpG sites in the *Reference DNA* tab, *Localization of CG dinucleotides* tab). In the next three columns, its coordinates are specified. Then, the position of the methylated cytosine (the C in forward, the G in reverse) is displayed, as well as its position in the raw and trimmed sequencing results. The position matching is obtained thanks to the alignment of sequences. For each position on the raw sequencing result, the peak height values of signals are extracted and used to compute the methylation percentage, with the following formulas, for forward and reverse sequencing respectively:

$$\text{Methylation percentage} = \frac{\text{peak}_C}{\text{peak}_C + \text{peak}_T} \times 100$$

$$\text{Methylation percentage} = \frac{\text{peak}_G}{\text{peak}_G + \text{peak}_A} \times 100$$

- **Combined** Methylation results from both sequencing results are then combined in a unique table with the calculation of the average methylation and standard deviation per position.
- **Individual methylation plot** Finally, a plot is generated to visualize results relative to the genomic sequence.

Output data

- **Directories** A diagram of output files directories is displayed (Figure 11).
- **Files** A list of all the output files with links to local folders is displayed.
- **Methylation data file preview** The output methylation data file which will be used as input for the grouped analysis is displayed as a table (The "*alg_coord_start*" and "*alg_coord_end*" columns contain a unique value corresponding to the start and end coordinates of the maximum aligned sequence for the individual analysis).

2.5.4 Output files

All the output files are located in the *results* directory, as depicted by the diagram in Figure 11.

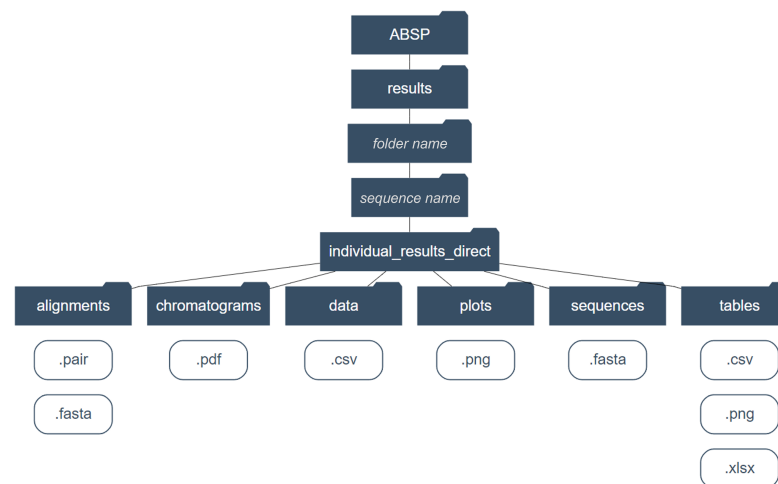


Figure 11. Diagram of output directories to locate output files from the individual analysis. The "*folder name*" and "*sequence name*" depend on the input entries when launching the analysis.

As for the same sequence (same primer set) results from both direct-BSP and cloning-BSP can be generated, the two types of outputs are separated into two subfolders: "*individual_results_direct*" and "*individual_results_cloning*".

- **alignments** In subfolders specific to each individual analysis, the .pair files of the alignments and the .fasta files of the aligned sequence are saved.
- **chromatograms** In subfolders specific to each individual analysis, the chromatograms of raw sequencing reads and the chromatograms of the trimmed sequencing are saved as .pdf files.
- **data** It contains all the methylation data .csv files for the grouped analysis.
- **plots** It contains all the individual methylation plots as well as legends for plots as .png image files.
- **sequences** The reference DNA sequences from plus and minus strands and the bisulfite converted sequence of the template strand are saved as .fasta files.
- **tables** In subfolders specific to each individual analysis, the data tables of raw sequencing results, the data tables of trimmed sequencing results, the summary table of sequencing trimming, the summary table of quality control, the bisulfite conversion rates table and the methylation percentages tables are saved as .csv .png and .xlsx files.

2.6 Grouped analysis

2.6.1 Procedure

In the **grouped analysis tab**, the left panel is to launch analysis, and the right panel provides entry information.

Experiment information

- **Select an existing folder.** Located in the "ABSP/results" folder, all of the analysis results will be saved in this folder. Having different folders of results can be used to separate the different analyses by projects, experiments, or users. Note that the six first letters of the folder name will appear in the report file name.
- **Select an existing sequence folder.** Located in the "ABSP/results/previous folder" folder, all of the analysis results will be saved in this folder for this sequence. This sequence name will be used in output files (tables, plots...) to refer to the sequence.
- **Select the reference genome.** It will only be used to display the genomic sequence in the genomic plot. See section 3.3 Code modifications at page 42 to add another genome in the drop-down list. The complete list of available genomes can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.
- **Select the experiment type.** The choice of the experiment can be either Direct-BSP or Cloning-BSP. The correct experiment type entry is essential to retrieve the methylation data files either in the "individual_results_direct" or "individual_results_cloning" folders.

Plot parameters

- **Select position labels for plots.** The CpG positions on plots can be referred to by different label types:
 - The **CpG coordinates** label type displays the genomic coordinates of the CpG site in the format *chr#:#####-#####* (e.g. "chr16:68771230-68771231").
 - The **CpG numbers** label type displays the CpG site identifier number on the represented sequence, from 1 to n.
 - The **None** label type displays blank labels, which can be a suitable alternative in case of extremely close CpG positions as labels may overlap.
- **Choose to separate plots by collections.** For lollipop plots, genomic plots, and boxplots and only for display purposes (it does not affect data). If this parameter is not ticked (default), all samples from different collections will be displayed on the same plot. If this parameter is ticked, a plot is generated by collection, displaying samples from one collection only.
- **Indicate the order of groups for display.** For this input to work, the folder,

sequence, and experiment type have to be selected and correct. In this case, the group names are extracted from methylation data files corresponding to the previously selected entries. The groups must all be selected, in the desired order for display.

- **Select the types of sample ordering for plots.** Four different sample ordering (ordinate axis ordering) are available for visualization plots, each provides a specific way of ordering samples on the ordinate axis based on different parameters. At least one must be selected, up to the four of them.
 - **As it is** arranges samples by alphabetic order of collections. If none or one collection is present, this order is equivalent to the *By groups* one.
 - **By groups** arranges samples by the provided group order above.
 - **By methylation levels** arranges samples depending on their methylation mean.
 - **By clusters** arranges samples depending on the hierarchical clustering calculated and represented by an associated dendrogram.

2.6.2 Output report

The HTML report *.html* file of the analysis is automatically saved in the "reports" folder in the ABSP directory.

Header First, in the top panel, the information about the sample experimental conditions is displayed in a table.

- Folder name
- Sequence name
- Reference genome
- Type of experiment
- Group order
- Date of analysis

Files content This tab summarizes all the data that have been used for this analysis.

- **Data files content** General information about the data is listed: sequence name, collections, groups, and replicates or clones.
- **Data files paths** Paths of the methylation files that were found and used for the analysis are listed.

Methylation data This tab regroups the methylation data in tables from the retrieved files.

- **Methylation data of replicates/clones** For each replicate or clone, depending on the experiment type, the methylation percentages of each CpG site are displayed in a table, with the mean of all positions and the associated standard deviation by replicate/clone.

For clones: Methylation percentages calculated from sequencing results are converted to 0% or 100% methylation status.

- A CpG site is considered unmethylated (0%) when the methylation percentage is between 0% and the defined threshold (default is 20%).
- A CpG site is considered methylated (100%) when the methylation percentage is between the defined threshold (default is 80%) and 100%.
- A CpG site partially methylated, with methylation percentage between 20% and 80%, is removed and annotated as NA (Not Available).
- For one clone, if more than a threshold percentage (default is 20%) of CpG sites are partially methylated, the clone is considered as defective and all of its CpG sites are annotated as NA (Not Available).

The thresholds can be modified in the script, please refer to section 3.3 Code modifications at page 42.

- **Methylation data of groups** For each group, the mean of methylation percentages of each CpG site are displayed in a table, with the mean of all positions and the associated standard deviation by groups.

Plots of replicates Only for direct-BSP. This tab provides plots to visualize the methylation data of replicates.

The Lollipop plots (condensed and proportional), the genomic plot, and the cluster dendrogram plot are as illustrated in Figure 12.

Plots of clones Only for cloning-BSP. This tab provides plots to visualize the methylation data of clones from each sample in separated plots.

Lollipop plots (condensed and proportional), genomic plot and cluster dendrogram plot are illustrated in Figure 13.

Plots of groups This tab provides plots to visualize the methylation data of groups per collection. For each sample, the mean of replicates or clones is calculated per CpG, and the results are displayed in this tab.

Lollipop plots (condensed and proportional), genomic plot and cluster dendrogram plot are illustrated in Figure 14.

Statistics This tab aims to compare methylation between groups, to find statistically significant differences, either by comparing CpG site per CpG site or by comparing the entire region.

- **Descriptive statistics of groups** Two tables of methylation data descriptive statistics are displayed: one for methylation data by CpG positions, and one for methylation data of all CpG positions for each sample.
- **Student's T test** Two tables display T-tests p-values between groups 2 by 2, one table for methylation data by CpG positions, and one for methylation data of all CpG positions for each sample.
- **Boxplots** Boxplots of methylation data with T tests p-values as numbers or symbols are generated, one plot details data of each CpG position (Figure 15

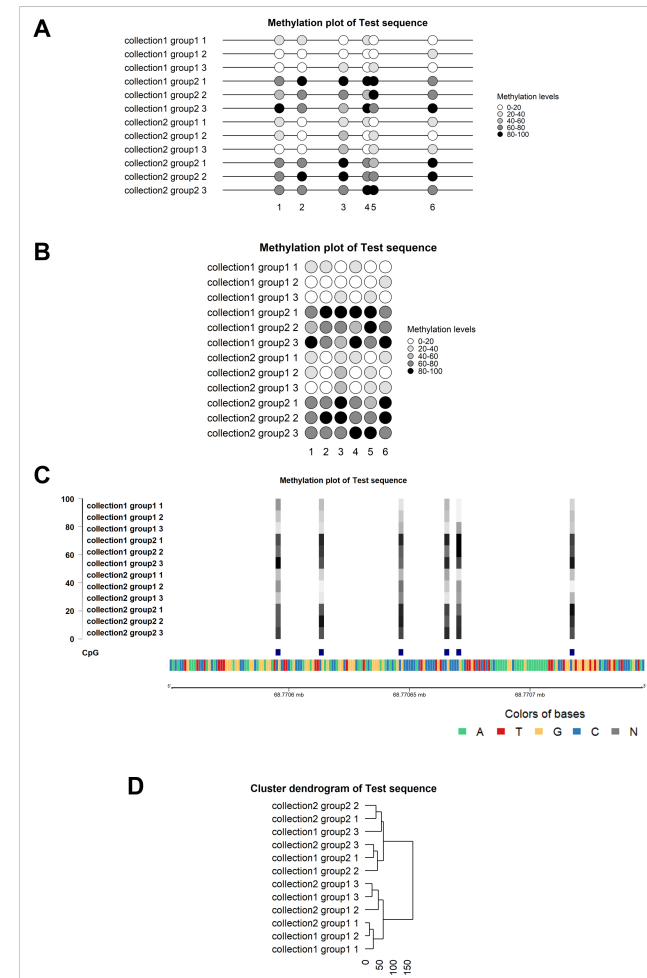


Figure 12. Visualization plots of all replicates (direct-BSP). The plots were generated based on mock methylation data for a test sequence. Three replicates (1, 2, and 3) per sample are represented on the plots, from two groups (group1 and group2) and two collections (collection1 and collection2) (sample ordering as *it is*). The methylation levels are given as percentages and correspond directly to the methylation output data from the individual analyses. **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering *by clusters*).

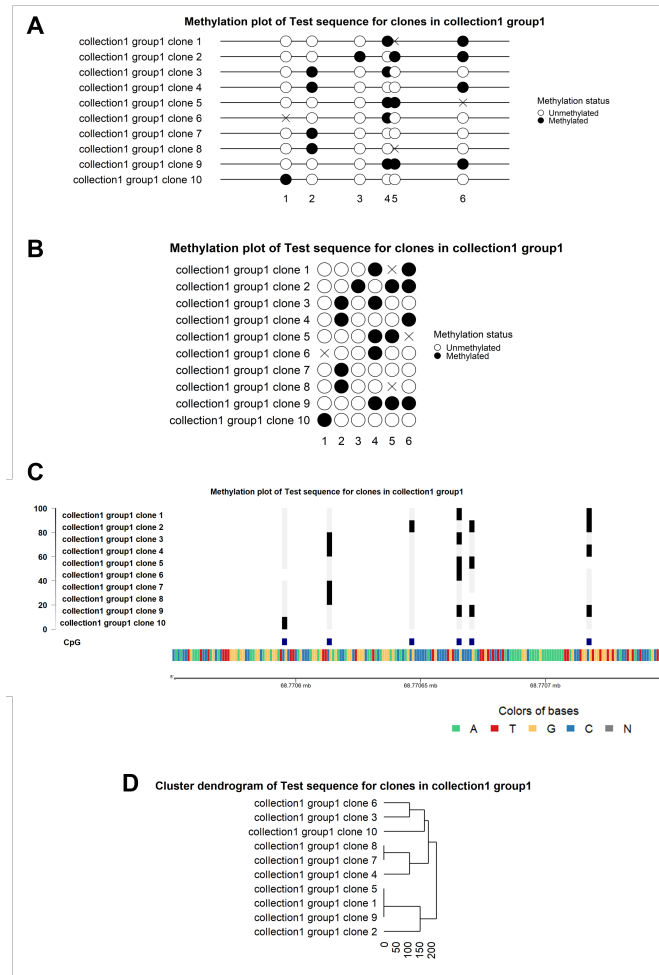


Figure 13. Visualization plots of all clones from one sample (cloning-BSP). The plots were generated based on mock methylation data for a test sequence. Ten clones (from 1 to 10) for the sample "collection1 group1" are represented on the plots (sample ordering as *it is*). The methylation status corresponds to the conversion of methylation percentages from the individual analyses into unmethylated (0%) or methylated (100%) (or not available) methylation statuses. **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering *by clusters*).

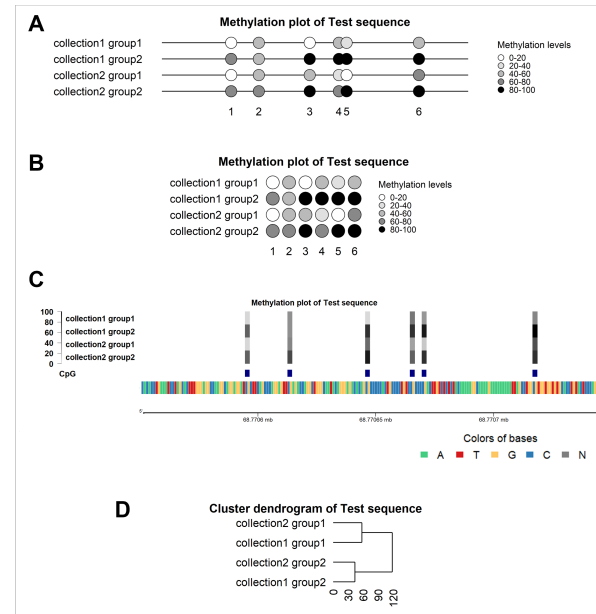


Figure 14. Visualization plots of groups (means of replicates/clones per sample). The plots were generated based on mock methylation data for a test sequence. Plots represent the means of methylation percentages of ten clones per sample and per CpG position (sample ordering as *it is*). **A.** Proportional lollipop plot. **B.** Condensed lollipop plot. **C.** Genomic plot. **D.** Cluster dendrogram (for sample ordering *by clusters*).

A for direct-BSP data and Figure 15 D for cloning-BSP), and another one data from methylation mean of the region (Figure 15 B for direct-BSP data and Figure 15 E for cloning-BSP).

- **Methylation profile plots** Methylation profile plots with Kruskal-Wallis tests p-values are generated for each collection (Figure 15 C and F).

The **Methylation plotter**^{7,8} tool served as a base to develop the different types of plots with different sample ordering.

⁷http://maplab.imppc.org/methylation_plotter/index.html

⁸Izaskun Mallona, Anna Díez-Villanueva, and Miguel A Peinado. "Methylation plotter: a web tool for dynamic visualization of DNA methylation data". In: *Source Code for Biology and Medicine* 9.1 (2014). Methylation plotter, p. 11. doi: [10.1186/1751-0473-9-11](https://doi.org/10.1186/1751-0473-9-11).

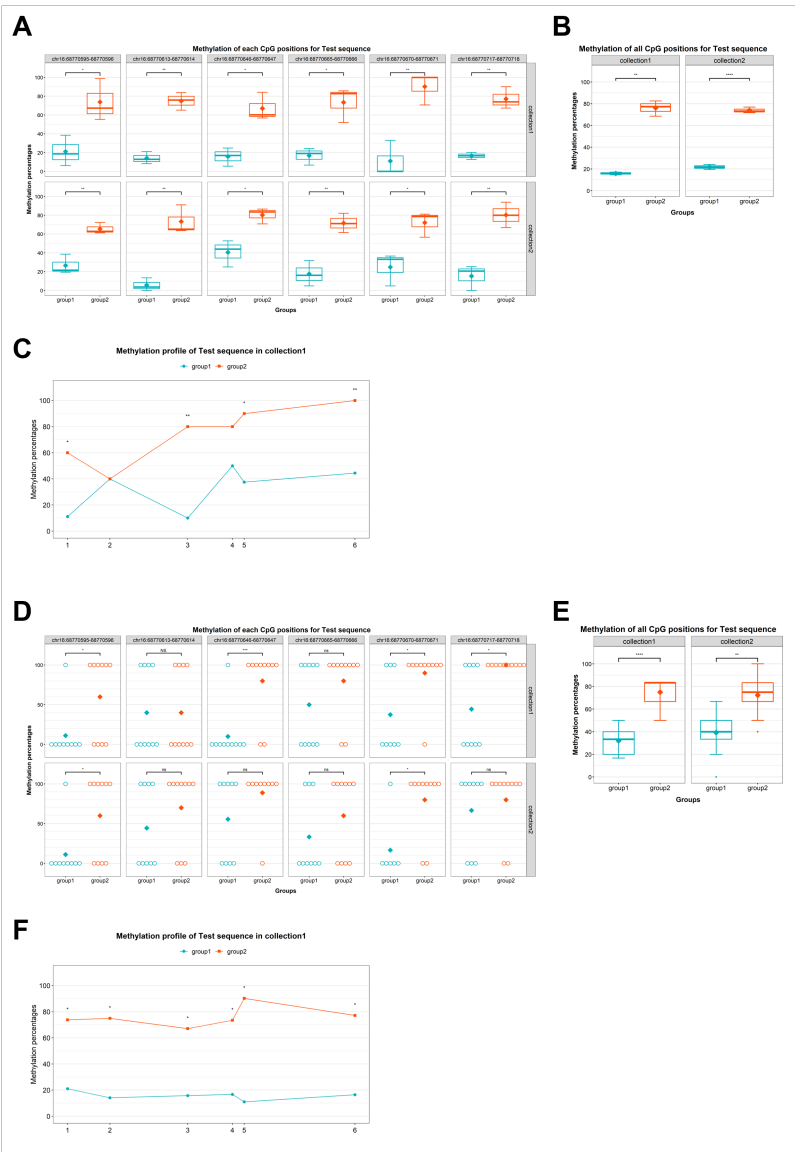


Figure 15. Boxplots and methylation profile plots. The plots were generated based on mock methylation data for a test sequence. **A.** Boxplots of direct-BSP methylation results for each CpG site. **B.** Boxplots of direct-BSP methylation results for means of all CpG sites. **C.** Methylation profile plot of direct-BSP methylation results for one of the two collections. **D.** Boxplots of cloning-BSP methylation results for each CpG site. As methylation levels of CpG from clones can only be either 0% or 100% thereby boxes can't be drawn, instead, each clone is represented in the plot by a circle. **E.** Boxplots of cloning-BSP methylation results for means of all CpG sites. **F.** Methylation profile plot of cloning-BSP methylation results for one of the two collections. In boxplots, symbols represent significance levels of Student's T-test p-values (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$). In methylation profile plots, symbols represent significance levels of Kruskal-Wallis test p-values (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$).

Output data

- **Directories** A diagram of output files directories is displayed (Figure 16).
- **Files** A list of all the output files with links to local folders is displayed.

2.6.3 Output files

All the output files are located in the *results* directory, as depicted by the diagram in Figure 16.

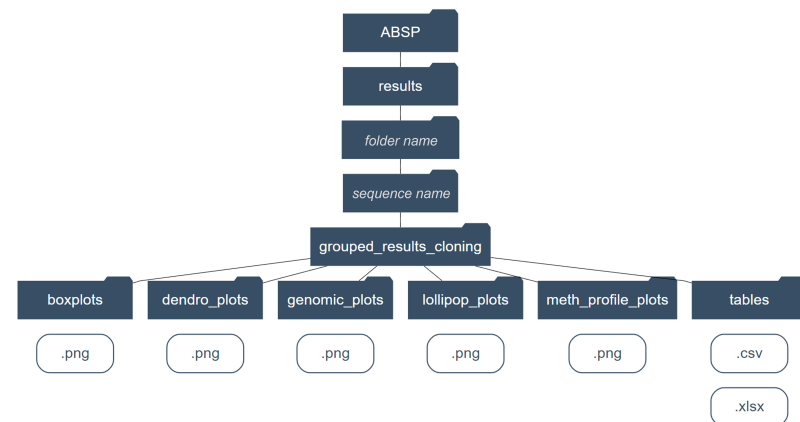


Figure 16. Diagram of output directories to locate output files from the individual analysis. The "folder name" and "sequence name" depend on the input entries when launching the analysis.

As results for the same sequence from both direct-BSP and cloning-BSP can be generated, the two types of outputs are separated into two subfolders: "grouped_results_dire

and "grouped_results_cloning".

- **boxplots** Boxplots are saved as .png files.
- **dendro_plots** Cluster dendrograms are saved as .png files.
- **genomic_plots** Genomic plots are saved as .png files, in subfolders for plots of replicates/clones and plots of groups.
- **lollipop_plots** Lollipop plots are saved as .png files, in subfolders for plots of replicates/clones and plots of groups.
- **meth_profile_plots** Methylation profile plots are saved as .png files
- **tables** The methylation data tables, the descriptive statistics tables of positions or means (means of all positions), and the Student's T test tables are saved as .csv and .xlsx files.

2.7 Multiple analyses

As described in the Figure 5, multiple analyses can be launched at the same time using data tables as input with all the required information. Both individual and grouped analyses can be launched, at the same time or separately.

2.7.1 Input files requirements

Two files are provided in the "documents" folder. They must be filled with the desired input entries, all information about how to fill the documents is also indicated within the documents as notes.

The documents are provided in the .xlsx (Microsoft Excel Open XML Format Spreadsheet) and .ods (OpenDocument Spreadsheet) formats, but the input file format must be either .xlsx or .csv (Comma-Separated Values) (.ods files must be converted to one of those formats).

- **Experiments data table for individual analyses:**
"multiple_individual_analyses_table.xlsx".
 - SEQUENCE NAME. The sequence name should be unique and consistent for each amplicon. It must not contain any special character.
 - COLLECTION. The collection refers to a separation of samples above the groups/conditions. Leave empty if you do not want to indicate a collection.
 - GROUP. The group refers to the condition that is to be compared. It must not contain any special character.
 - REPLICATE NUMBER. The replicate number refers to the experiment repetition identifier. It must be an integer ≥ 1 . Only for direct-BSP experiments, leave empty otherwise.
 - CLONE NUMBER. The clone number refers to the individual clone identifier. It must be an integer ≥ 1 . Only for cloning-BSP experiments, leave empty otherwise.

- GENOME. The genome refers to the reference genome assembly used for coordinates and plots displaying the genomic sequence. Only a short list of genomes is displayed in the cells, but you can use another available genome. Please refer to the provided document: "List of BSgenomes" to get the list of available genomes assemblies and the correct spelling.
 - PATH TO FASTA FILE OF REFERENCE DNA. Path to the .fasta file of the reference DNA sequence on your computer. On Windows, you can copy the file path by holding down shift then right-clicking on the file, and selecting "Copy as path" in the menu. On macOS, you can copy the file path by right-clicking on the file to display the menu then holding down the option key and selecting "Copy ... as Pathname".
 - DATE SEQUENCING #1. Date of the sequencing result #1 for traceability. The date format must be YYYY-MM-DD and the cell format must be set to "date".
 - PATH TO SEQUENCING FILE #1. Path to sequencing result #1 .ab1 file.
 - DATE OF SEQUENCING #2. Date of the sequencing result #2 for traceability. The date format must be YYYY-MM-DD and the cell format must be set to "date".
 - PATH TO SEQUENCING FILE #2. Path to sequencing result #2 .ab1 file.
- **Parameters table for grouped analyses:**
"multiple_grouped_analyses_table.xlsx".
 - SEQUENCE NAME. The sequence name should be unique and consistent for each amplicon. It must not contain any special character.
 - GENOME. The genome refers to the reference genome assembly used for coordinates and plots displaying the genomic sequence. Only a short list of genomes is displayed in the cells, but you can use another available genome. Please refer to the provided document: "List of BSgenomes" to get the list of available genomes assemblies and the correct spelling.
 - EXPERIMENT TYPE. The experiment type can be either Direct-BSP or Cloning-BSP.
 - CpG POSITION LABEL TYPE. The CpG position label type refers to the displayed element corresponding to the CpG position on plots.
 - SEPARATION OF COLLECTIONS. This parameter refers to the generation of plots: if FALSE, all samples from all collections will be displayed on the same plot, and if TRUE, samples will be split into different plots, one plot per collection.
 - LIST OF GROUPS ORDERED. The cell must contain all of the groups within the experiment in the order you want them to be displayed. Type your groups separated by commas. It must not contain any special character.
 - TYPE OF SAMPLE ORDERING. The type of sample ordering refers to the parameter used for ordering samples on plots. Four types of ordering are available: As it is, By groups, By methylation levels, and By clusters. Multiple ordering can be chosen by typing the ordering names separated by commas.

2.7.2 Procedure

- **Select an existing folder** within the ABSP results folder to locate all of the analysis results. To create a new folder, select the "Create new folder" entry and enter the name of the new folder in the text input. Note that the six first letters will appear in the report file name.
- **Select the filled table as input.** Both the experiments data table and the grouped parameters table can be provided at the same time to launch individual analyses followed by grouped analyses, or only one of the two tables can be provided and will launch the corresponding analyses, either individual analyses or grouped analyses.
- **Launch the analyses** by clicking on the bottom button "Run analyses".

2.7.3 Output files

The reports and output files of analyses are saved in the *reports* and *results* folders, in the same way as for the manual launch of individual and grouped analyses, see sections 2.5 Individual analysis at page 19 and 2.6 Grouped analysis at page 29.

3 Complementary information

3.1 Some recommendations for the BSP experiment

The length of the PCR products should not exceed 350-400 bp, as the bisulfite treatment causes DNA strand breakages long amplicon can not be properly amplified, and ABSP plot display for sequences above 400 bp with numerous CpG is not optimal.

Several tools can be used to design BSP primers, several are listed in the [Methtools](#)⁹ list of tools, such as [Methprimer](#)¹⁰ to design primers and [BiSearch](#)¹¹ to check for unintended PCR products, on bisulfite-treated DNA.

3.2 Detailed workflow of ABSP individual analysis

In [Figure 17](#), the individual analysis input, steps and outputs are described in more detail.

⁹<http://bigd.big.ac.cn/methbank/methTool/list/>

¹⁰<https://www.urogene.org/methprimer/>

¹¹<http://bisearch.enzim.hu/>

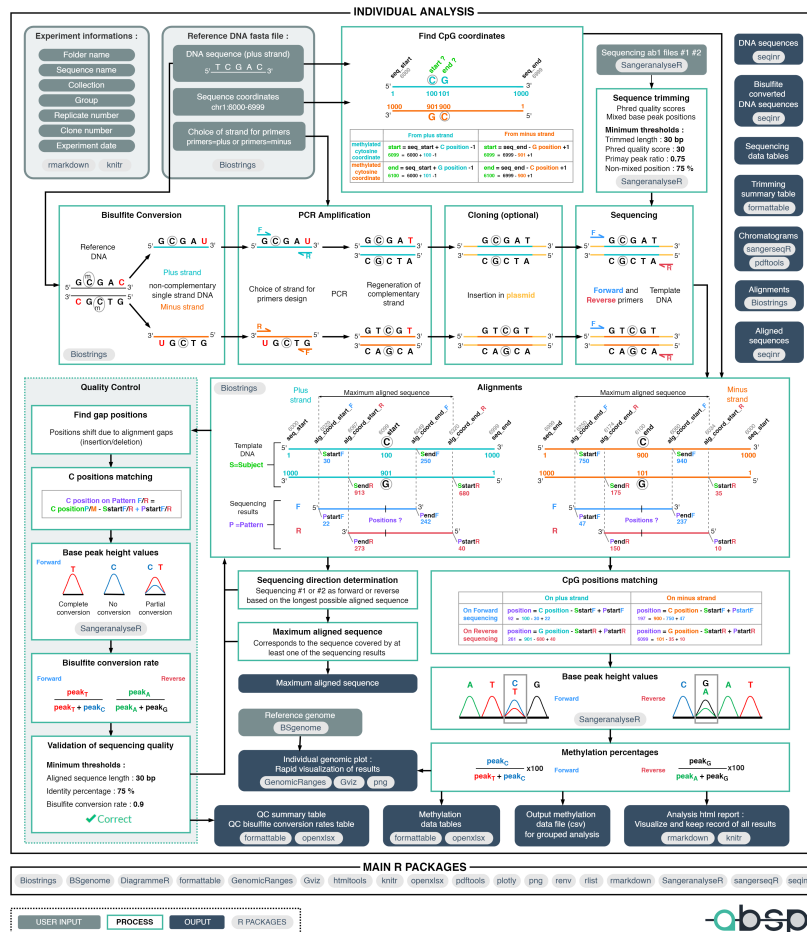


Figure 17. Detailed workflow of the individual analysis.

3.3 Code modifications

3.3.1 List of reference genomes

In the individual analysis and grouped analysis tabs of the app, the drop-down list to select the reference genome is limited. If your reference genome does not appear it doesn't mean that it is not available, and can be manually added. The complete list of [genomes assemblies](#)¹² with the correct spelling can be found in the "List of BSgenomes.xlsx" file in the "ABSP/documents" folder.

To modify the displayed drop-down list items, the *app.R* script has to be modified. In the [Code listing 1](#), the "list_genomes" object corresponds to the vector of character strings listing the displayed reference genome. Any of the other reference genomes can be added to the list.

[Code listing 1](#). List of reference genomes displayed in the drop-down lists as selectable inputs for analyses. From the "app.R" script.

```

14 # Here, to simplify, a short list of genomes is displayed but all BSgenome
    can be used
15 # To get the list of all genomes of BSgenome package run : 'BSgenome::
    available.genomes()', more information on genomes at https://genome.
    ucsc.edu/cgi-bin/hgGateway
16 # A new genome can be added to the list displayed just below :
17 list_genomes <- c(
    "BSgenome.Hsapiens.UCSC.hg19", "BSgenome.Hsapiens.UCSC.hg38", "
    BSgenome.Mmusculus.UCSC.mm10", "BSgenome.Mmusculus.UCSC.mm39", "
    BSgenome.Rnorvegicus.UCSC.rn6", "BSgenome.Rnorvegicus.UCSC.rn7", "
    BSgenome.Cfamiliaris.UCSC.canFam3", "BSgenome.Mmulatta.UCSC.rheMac8",
    "BSgenome.Ggallus.UCSC.galGal6", "BSgenome.Drerio.UCSC.danRer11", "
    BSgenome.Celegans.UCSC.ce11", "BSgenome.Dmelanogaster.UCSC.dm6")

```

To modify the genome name selected by default in the drop-down "Select genome" list of either the individual or grouped analysis tabs, change the name of the genome in the following lines ([Code listing 2](#) and [Code listing 3](#)):

[Code listing 2](#). Genome name selected by default in the drop-down list of the individual analysis tab. From the "app.R" script.

```

249 selectInput("genomeI", label = "Select genome", choices=list_genomes, selected
    ="BSgenome.Hsapiens.UCSC.hg19")

```

3.3.2 Modify the default thresholds

¹²<https://genome.ucsc.edu/cgi-bin/hgGateway>

Code listing 3. Genome name selected by default in the drop-down-list of the grouped analysis tab. From the *"app.R"* script.

```
524 selectInput("genomeG", label = "Select genome", choices=list_genomes, selected
      = "BSgenome.Hsapiens.UCSC.hg19"),
```

Individual analysis

The default thresholds used for the individual analysis can be modified ([Code listing 4](#)).

Code listing 4. Default thresholds in the individual analysis script. From the *"ABSP_individual_analysis.Rmd"* script.

```
602 '''{r Thresholds, include=F}
603
604 # Thresholds
605
606 # Maximum base-calling error probability (value per position):
607 th_quality_error <- 0.001
608 # Minimum phred quality score, logarithmically linked to error probability
609   (value per position):
610 th_quality_phred <- (-10*log(th_quality_error, 10))
611
612 # Minimum ratio of primary peak, corresponding to the primary peak value
613   over the total of peak value, to consider a position as non-mixed (
614   value per position):
615 th_mixed_position <- 0.75
616 # Minimum percentage of non-mixed positions in the trimmed sequence to be
617   considered as non-mixed (value for the total trimmed sequence):
618 th_mixed_perc <- 75 # %
619
620 # Minimum length of trimmed sequences
621 th_min_trim <- 30 # bp
622
623 # Minimum length of aligned sequences
624 th_min_alg <- 30 # bp
625 '''
```

Grouped analysis

The default thresholds used for the grouped analysis, for clones methylation percentage conversion into methylation status (0% or 100%), can be modified ([Code listing 5](#)).

3.3.3 Modify the plots colors and point shapes

For the grouped analysis, the plot colors and point shape parameters can be modified. As depicted in [Code listing 6](#), the color of bases can be changed in the

Code listing 5. Default thresholds in the grouped analysis script. From the *"ABSP_grouped_analysis.Rmd"* script.

```
524 '''{r Thresholds, include=F}
525
526 # Thresholds (used for cloning only) :
527
528 # unmethylated clones : methylation between 0% and 20%
529 th_unmethylated_max <- 20
530
531 # methylated clones : methylation between 80% and 100%
532 th_methylated_min <- 80
533
534 # maximum proportion of partial positions allowed : 20% of CpG positions
535 th_partialpos_ratio <- 0.2
536
537 clone_thresholds <- c(th_unmethylated_max, th_methylated_min, th_
538   partialpos_ratio)
539 '''
```

bases_colors object, the plot colors for each group in the *plot_colors* object, and the point shapes for each group in the *plot_shapes* object.

In the *ABSP_grouped_analysis.Rmd* script, the code chunk below the one depicted in the [Code listing 6](#) gives examples of different palette colors and points shapes, with functions to visualize them.

Code listing 6. Colors and shapes setting for plots in the grouped analysis script. From the *ABSP_grouped_analysis.Rmd* script.

```

359 ''{r colors and shapes settings, include=F}
360 # Colors of sequence track for genomic plots
361 bases_colors <- c(A="#43CD80", T="#D7191C", G="#FDC661", C="#2C7BB6", N="#
    7F7F7F")
362
363 # Colors of groups for plots
364 plot_colors <- c("aqua"="#00AFBB",
365                 "tangerine"="#FC4E07",
366                 "sun"="#E7B800",
367                 "berry"="#d30446",
368                 "lime"="#90c613",
369                 "grape"="#7839de",
370                 "flamingo"="#d12a97",
371                 "jade"="#00b673",
372                 "ink"="#1221ed",
373                 "terracotta"="#a93b2c")
374
375 # Shapes of groups for methylation profile plots
376 plot_shapes <- c("round"=19,
377                 "square"=15,
378                 "triangle"=17,
379                 "round_border"=21,
380                 "square_border"=22,
381                 "triangle_border"=24,
382                 "diamond"=18,
383                 "small_round"=20,
384                 "reverse_triangle_border"=25,
385                 "diamond_border"=23)
386 ''

```

4 Troubleshooting guide

4.1 General

Error	Cause	Solution
After following the opening procedure (for the first time) and clicking on the "Run App" button, the app interface is not opening.	Package installation failed or some packages are not fully installed. Plus, to finalize some package installation, R needs to be restarted.	In the RStudio console, type "renv::restore(prompt=F)" and press enter to run it. The download and installation of packages should start. Then, in the top bar, click on "Session" and "Restart R". Then try to open the app by clicking on the "Run app" button. If the problem is not solved, please contact us with a description of the issue.
Warning message about RStudio version in the RStudio console. <i>"R graphics engine version X is not supported by this version of RStudio. The Plots tab will be disabled until a newer version of RStudio is installed."</i>	The RStudio installed on the device is not up-to-date.	Download and install the newest version of RStudio at https://www.rstudio.com/products/rstudio/download/ .
Prompt message displayed in the RStudio console during the selected BSgenome installation. <i>"Do you want to install from sources the package which needs compilation? (Yes/no/cancel)"</i>	The installation of the BSgenome requires other packages to be installed.	Type "Yes" in the console panel and press enter to run it.

<p>Warning message when opening the ABSP about project configuration and R version (and the app does not function). <i>"This project is configured to use R version 'X.X.X', but 'X.X.X' is currently being used."</i></p>	<p>The R version does not match the R project configuration, the R version is either older or newer than the one used to generate the ABSP project.</p>	<p>Either download the latest version of R to match the one of the project configuration (https://www.r-project.org/) or download the latest version of ABSP to match the R version (https://github.com/ABSP-methylation-tool/ABSP). If the currently available ABSP version does not match the currently available R version, download the corresponding version of R and contact us, for us to update ABSP as soon as possible to match the current R version.</p>
--	---	---

4.2 Individual analysis

Error	Cause	Solution
<p>Analysis report aborted after the reference DNA step. <i>Error: Length of reference DNA sequence does not match with the provided genomic coordinates. Please verify concordance between the reference DNA sequence and genomic coordinates.</i></p>	<p>The provided reference sequence in the fasta file has not had the same length as the one calculated based on genomic coordinates provided in the fasta file header. So the sequence and coordinates might not match.</p>	<p>Check the reference and its coordinates. If the IGV (Integrative Genomics Viewer) software is used to get them, make sure to properly verify the first and last nucleotides, as there can be a one nucleotide difference between the coordinates and the actual sequence.</p>
<p>Analysis report aborted after the sequencing trimming step. <i>Error: Analysis has been stopped as none of the sequencing results are of sufficient quality to be used.</i></p>	<p>The provided sequencing results are not of good quality, the trimming step was not able to find a trimmed sequence long enough (below length threshold) and/or with poor quality (below the quality score and/or non-mixed positions thresholds).</p>	<p>Lowering a bit the trimming thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>
<p>Analysis report aborted after the alignment step. <i>Error: Analysis has been stopped as none of the possible alignments are of sufficient length (< N bp) to be used.</i></p>	<p>Even if the sequencing results were of good quality and successfully trimmed, their alignments with the reference DNA sequence give too short aligned sequences to pursue the analysis.</p>	<p>Check if the reference DNA sequence is the correct one. Check if the sequencing results passed the trimming step with values just above thresholds that might explain the alignment results. In this case, lowering some thresholds may solve the issue. If not, the best option is to provide new results from a new sequencing run.</p>

<p>Analysis report aborted after the alignment step. <i>Error: Analysis has been stopped as sequencing result direction (forward or reverse) could not be found.</i></p>	<p>This error occurs when the determined direction for both sequencing results happened to be the same, or when only one sequencing has been successfully trimmed and aligned, but alignments as forward or reverse have the same length.</p>	<p>For the first case, the results might have passed the alignment step with values just below thresholds, explaining the identical length of the aligned sequence in both directions, still, the best option should be to provide new sequencing results to have better trimmed and aligned results. For the second case, lowering a bit some thresholds might help the other sequencing to pass the trimming and alignment steps, still, the best option should be to provide new sequencing results to have better trimmed and aligned results.</p>
<p>Analysis report aborted after the alignment step. <i>Error: Analysis has been stopped as none of the provided sequencing reads are of sufficient quality to be used.</i></p>	<p>The provided sequencing results are not of good quality, they passed the trimming step but not the alignment step. The aligned sequences were not long enough (below length threshold) or direction determination failed.</p>	<p>Lowering a bit the trimming and alignment thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>
<p>Analysis report aborted after the Quality Control (QC) step. <i>Error: Analysis has been stopped as sequencing results are defined as incorrect by Quality Control. The analysis can not be performed.</i></p>	<p>The provided sequencing results are not of good quality when compared to the reference DNA; they passed the trimming and alignment steps but the quality control found the sequencing results incorrect: either the identity percentage or the mean bisulfite conversion rates are below the threshold.</p>	<p>Lowering a bit the thresholds might solve the issue, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>

<p>Analysis report aborted after the Quality Control (QC) step. <i>Error: Analysis has been stopped as no CpG sites were found covered by sequencing results.</i></p>	<p>Even if the sequencing results passed the trimming, alignment, and quality control steps, the aligned results are not long enough and do not cover any CpG sites on the sequence, thereby methylation levels can not be computed.</p>	<p>Lowering a bit the thresholds might solve the issue to get longer aligned sequences, but if not, no solution can be provided. The best recommendation is to perform the sequencing run one more time. If the sequencing is not of good quality again, then there must be an experimental issue with BSP samples or sequencing runs.</p>
<p>Analysis failed and no report has been generated</p>	<p>Unexpected error.</p>	<p>Please contact us and send us the error message appearing in the RStudio console.</p>

4.3 Grouped analysis

Error	Cause	Solution
<p>Analysis report aborted after the header. <i>Error: No file methylation data files found. Check inputs: folder name and sequence name, and check the 'data' directory in the individual results folder.</i></p>	<p>No methylation data files from the individual analysis, corresponding to the input information, were able to be retrieved.</p>	<p>Check if the individual analyses have already been run, if the files were not moved to another folder, they should be in the 'data' directory in the individual results folder, or if the input information, folder name, sequence name, and experiment type are correct. If the issue is not solved by these recommendations, please contact us.</p>
<p>Analysis failed and no report has been generated.</p>	<p>Unexpected error.</p>	<p>Please contact us and send us the error message appearing in the RStudio console.</p>

5 Acknowledgments

An important part of the plot generation work for the grouped analysis was based on the R code provided by the [Methylation plotter](#)^{13,14} tool.

Main R packages used:

- arrangements
- BiocManager
- Biostrings
- BSgenome
- compareGroups
- DiagrammeR
- dplyr
- formattable
- GenomelInfoDb
- GenomicRanges
- ggdendro
- ggplot2
- ggpubr
- Gviz
- htmltools
- htmlwidgets
- knitr
- openxlsx
- pdftools
- plotly
- png
- purrr
- RColorBrewer
- readr
- renv
- rlist
- rmarkdown
- Rmisc
- rstatix
- sangeranalyseR
- sangerseqR
- seqinr
- shiny
- shinythemes
- webshot

¹³http://maplab.imppc.org/methylation_plotter/index.html

¹⁴Mallona, Díez-Villanueva, and Peinado, "Methylation plotter: a web tool for dynamic visualization of DNA methylation data".

List of all R packages used:

- **abind 1.4.5** - Plate T, Heiberger R (2016). `_abind: Combine Multidimensional Arrays_`. R package version 1.4-5, <https://CRAN.R-project.org/package=abind>.
- **ade4 1.7.19** - Dray S, Dufour A (2007). "The ade4 Package: Implementing the Duality Diagram for Ecologists." *Journal of Statistical Software*, *22*(4), 1-20. doi:10.18637/jss.v022.i04 <https://doi.org/10.18637/jss.v022.i04>.
- **AnnotationDbi 1.58.0** - Pagès H, Carlson M, Falcon S, Li N (2022). `_AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor_`. R package version 1.58.0, <https://bioconductor.org/packages/AnnotationDbi>.
- **AnnotationFilter 1.20.0** - Morgan M, Rainer J (2022). `_AnnotationFilter: Facilities for Filtering Bioconductor Annotation Resources_`. R package version 1.20.0, <https://github.com/Bioconductor/AnnotationFilter>.
- **anytime 0.3.9** - Eddelbuettel D (2020). `_anytime: Anything to 'POSIXct' or 'Date' Converter_`. R package version 0.3.9, <https://CRAN.R-project.org/package=anytime>.
- **arrangements 1.1.9** - Lai R (2020). `_arrangements: Fast Generators and Iterators for Permutations, Combinations, Integer Partitions and Compositions_`. R package version 1.1.9, <https://CRAN.R-project.org/package=arrangements>.
- **askpass 1.1** - Ooms J (2019). `_askpass: Safe Password Entry for R, Git, and SSH_`. R package version 1.1, <https://CRAN.R-project.org/package=askpass>.
- **assertthat 0.2.1** - Wickham H (2019). `_assertthat: Easy Pre and Post Assertions_`. R package version 0.2.1, <https://CRAN.R-project.org/package=assertthat>.
- **backports 1.4.1** - Lang M, R Core Team (2021). `_backports: Reimplementations of Functions Introduced Since R-3.0.0_`. R package version 1.4.1, <https://CRAN.R-project.org/package=backports>.
- **base64enc 0.1.3** - Urbaneck S (2015). `_base64enc: Tools for base64 encoding_`. R package version 0.1-3, <https://CRAN.R-project.org/package=base64enc>.
- **bbmle 1.0.25** - Bolker B, R Development Core Team (2022). `_bbmle: Tools for General Maximum Likelihood Estimation_`. R package version 1.0.25, <https://CRAN.R-project.org/package=bbmle>.
- **bdsmatrix 1.3.6** - Therneau T (2022). `_bdsmatrix: Routines for Block Diagonal Symmetric Matrices_`. R package version 1.3-6, <https://CRAN.R-project.org/package=bdsmatrix>.
- **beachmat 2.12.0** - Lun ATL, Pagès H, Smith ML (2018). "beachmat: A Bioconductor C++ API for accessing high-throughput biological data from a variety of R matrix types." *PLoS Comput. Biol.*, *14*(5), e1006135. doi:10.1371/journal.pcbi.1006135 <https://doi.org/10.1371/journal.pcbi.1006135>.
- **BH 1.78.0.0** - Eddelbuettel D, Emerson JW, Kane MJ (2021). `_BH: Boost C++ Header Files_`. R package version 1.78.0-0, <https://CRAN.R-project.org/package=BH>.
- **Biobase 2.56.0** - Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Ole's AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015). "Orchestrating high-throughput genomic analysis with Bioconductor." *Nature Methods*, *12*(2), 115-121. <http://www.nature.com/nmeth/journal/v12/n2/full/nmeth.3252.html>.
- **BiocGenerics 0.42.0** - Huber, W., Carey, J. V, Gentleman, R., Anders, S., Carlson, M., Carvalho, S. B, Bravo, C. H, Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, D. K, Irizarry, A. R, Lawrence, M., Love, I. M, MacDonald, J., Obenchain, V., Ole's, K. A, Pag'es, H., Reyes, A., Shannon, P, Smyth, K. G, Tenenbaum, D., Waldron, L., Morgan, M. (2015). "Orchestrating high-throughput genomic analysis with Bioconductor." *Nature Methods*, *12*(2), 115-121. <http://www.nature.com/nmeth/journal/v12/n2/full/nmeth.3252.html>.
- **BiocManager 1.30.18** - Morgan M (2022). `_BiocManager: Access the Bioconductor Project Package Repository_`. R package version 1.30.18, <https://CRAN.R-project.org/package=BiocManager>.
- **BiocParallel 1.30.3** - Morgan M, Wang J, Obenchain V, Lang M, Thompson R, Turaga N (2022). `_BiocParallel: Bioconductor facilities for parallel evaluation_`. R package version 1.30.3, <https://github.com/Bioconductor/BiocParallel>.
- **BiocStyle 2.24.0** - Ole's A (2022). `_BiocStyle: Standard styles for vignettes and other Bioconductor documents_`. R package version 2.24.0, <https://github.com/Bioconductor/BiocStyle>.
- **Biostrings 2.64.0** - Pagès H, Aboyoun P, Gentleman R, DebRoy S (2022). `_Biostrings: Efficient manipulation of biological strings_`. R package version 2.64.0, <https://bioconductor.org/packages/Biostrings>.
- **bit 4.0.4** - Oehlschlägel J, Ripley B (2020). `_bit: Classes and Methods for Fast Memory-Efficient Boolean Selections_`. R package version 4.0.4, <https://CRAN.R-project.org/package=bit>.
- **bit64 4.0.5** - Oehlschlägel J, Silvestri L (2020). `_bit64: A S3 Class for Vectors of 64bit Integers_`. R package version 4.0.5, <https://CRAN.R-project.org/package=bit64>.
- **bitops 1.0.7** - port SobSDiR, revised ebMM, Dutky mbS (2021). `_bitops: Bitwise Operations_`. R package version 1.0-7, <https://CRAN.R-project.org/package=bitops>.
- **blob 1.2.3** - Wickham H (2022). `_blob: A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')_`. R package version 1.2.3, <https://CRAN.R-project.org/package=blob>.
- **bookdown 0.29** - Xie Y (2022). `_bookdown: Authoring Books and Technical Documents with R Markdown_`. R package version 0.29, <https://github.com/rstudio/bookdown>.
- **brio 1.1.3** - Hester J, Csárdi G (2021). `_brio: Basic R Input Output_`. R package version 1.1.3, <https://CRAN.R-project.org/package=brio>.
- **broom 1.0.1** - Robinson D, Hayes A, Couch S (2022). `_broom: Convert Statistical Objects into Tidy Tibbles_`. R package version 1.0.1, <https://CRAN.R-project.org/package=broom>.
- **BSgenome 1.64.0** - Pagès H (2022). `_BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs_`. R package version 1.64.0, <https://bioconductor.org/packages/BSgenome>.
- **bslib 0.4.0** - Sievert C, Cheng J (2022). `_bslib: Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'_`. R package version 0.4.0, <https://CRAN.R-project.org/package=bslib>.
- **bsseq 1.32.0** - Hansen KD, Langmead B, Irizarry RA (2012). "BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions." *Genome Biology*, *13*(10), R83. doi:10.1186/gb-2012-13-10-r83 <https://doi.org/10.1186/gb-2012-13-10-r83>.
- **cachem 1.0.6** - Chang W (2021). `_cachem: Cache R Objects with Automatic Pruning_`. R package version 1.0.6, <https://CRAN.R-project.org/package=cachem>.
- **callr 3.7.2** - Csárdi G, Chang W (2022). `_callr: Call R from R_`. R package version 3.7.2, <https://CRAN.R-project.org/package=callr>.
- **car 3.1.0** - Fox J, Weisberg S (2019). `_An R Companion to Applied Regression_`, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/>

Companion/.

- **carData 3.0.5** - Fox J, Weisberg S, Price B (2022). `_carData`: Companion to Applied Regression Data Sets. R package version 3.0-5, <https://CRAN.R-project.org/package=carData>.
- **caTools 1.18.2** - Tuszynski J (2021). `_caTools`: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc. R package version 1.18.2, <https://CRAN.R-project.org/package=caTools>.
- **cellranger 1.1.0** - Bryan J (2016). `_cellranger`: Translate Spreadsheet Cell Ranges to Rows and Columns. R package version 1.1.0, <https://CRAN.R-project.org/package=cellranger>.
- **checkmate 2.1.0** - Lang M (2017). "checkmate: Fast Argument Checks for Defensive R Programming." *The R Journal*, *9*(1), 437-445. doi:10.32614/RJ-2017-028 <https://doi.org/10.32614/RJ-2017-028>.
- **ChIPpeakAnno 3.30.1** - Zhu L, Gazin C, Lawson N, Pagès H, Lin S, Lapointe D, Green M (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data." *BMC Bioinformatics*, *11*(1), 237. ISSN 1471-2105, doi:10.1186/1471-2105-11-237 <https://doi.org/10.1186/1471-2105-11-237>, <http://www.biomedcentral.com/1471-2105/11/237>.
- **chron 2.3.57** - James D, Hornik K (2022). `_chron`: Chronological Objects which Can Handle Dates and Times. R package version 2.3-57. S original by David James, R port by Kurt Hornik, <https://CRAN.R-project.org/package=chron>.
- **cli 3.3.0** - Csárdi G (2022). `_cli`: Helpers for Developing Command Line Interfaces. R package version 3.3.0, <https://CRAN.R-project.org/package=cli>.
- **clipr 0.8.0** - Lincoln M (2022). `_clipr`: Read and Write from the System Clipboard. R package version 0.8.0, <https://CRAN.R-project.org/package=clipr>.
- **cluster 2.1.4** - Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022). `_cluster`: Cluster Analysis Basics and Extensions. R package version 2.1.4 - For new features, see the 'Changelog' file (in the package source), <https://CRAN.R-project.org/package=cluster>.
- **coda 0.19.4** - Plummer M, Best N, Cowles K, Vines K (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, *6*(1), 7-11. <https://journal.r-project.org/archive/>.
- **colorspace 2.0.3** - Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrell P, Stauffer R, Wilke CO (2020). "colorspace: A Toolbox for Manipulating and Assessing Colors and Palettes." *Journal of Statistical Software*, *96*(1), 1-49. doi:10.18637/jss.v096.i01 <https://doi.org/10.18637/jss.v096.i01>.
- **commonmark 1.8.0** - Ooms J (2022). `_commonmark`: High Performance CommonMark and Github Markdown Rendering in R. R package version 1.8.0, <https://CRAN.R-project.org/package=commonmark>.
- **compareGroups 4.5.1** - Subirana I, Sanz H, Vila J (2014). "Building Bivariate Tables: The compareGroups Package for R." *Journal of Statistical Software*, *57*(12), 1-16. <https://www.jstatsoft.org/v57/i12/>.
- **corrplot 0.92** - Wei T, Simko V (2021). `_R package 'corrplot'`: Visualization of a Correlation Matrix. (Version 0.92), <https://github.com/taiyun/corrplot>.
- **cowplot 1.1.1** - Wilke C (2020). `_cowplot`: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1, <https://CRAN.R-project.org/package=cowplot>.
- **cpp11 0.4.2** - Hester J, François R (2021). `_cpp11`: A C++11 Interface for R's C Interface. R package version 0.4.2, <https://CRAN.R-project.org/package=cpp11>.
- **crayon 1.5.1** - Csárdi G (2022). `_crayon`: Colored Terminal Output. R package version 1.5.1, <https://CRAN.R-project.org/package=crayon>.
- **crosstalk 1.2.0** - Cheng J, Sievert C (2021). `_crosstalk`: Inter-Widget Interactivity for HTML Widgets. R package version 1.2.0, <https://CRAN.R-project.org/package=crosstalk>.
- **curl 4.3.2** - Ooms J (2021). `_curl`: A Modern and Flexible Web Client for R. R package version 4.3.2, <https://CRAN.R-project.org/package=curl>.
- **data.table 1.14.2** - Dowle M, Srinivasan A (2021). `_data.table`: Extension of 'data.frame'. R package version 1.14.2, <https://CRAN.R-project.org/package=data.table>.
- **DBI 1.1.3** - R Special Interest Group on Databases (R-SIG-DB), Wickham H, Müller K (2022). `_DBI`: R Database Interface. R package version 1.1.3, <https://CRAN.R-project.org/package=DBI>.
- **dbplyr 2.2.1** - Wickham H, Girlich M, Ruiz E (2022). `_dbplyr`: A 'dplyr' Back End for Databases. R package version 2.2.1, <https://CRAN.R-project.org/package=dbplyr>.
- **DelayedArray 0.22.0** - Pagès H, Hickey wcfP, Lun A (2022). `_DelayedArray`: A unified framework for working transparently with on-disk and in-memory array-like datasets. R package version 0.22.0, <https://bioconductor.org/packages/DelayedArray>.
- **DelayedMatrixStats 1.18.0** - Hickey P (2022). `_DelayedMatrixStats`: Functions that Apply to Rows and Columns of 'DelayedMatrix' Objects. R package version 1.18.0, <https://github.com/PeteHaitch/DelayedMatrixStats>.
- **deldir 1.0.6** - Turner R (2021). `_deldir`: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation. R package version 1.0-6, <https://CRAN.R-project.org/package=deldir>.
- **desc 1.4.2** - Csárdi G, Müller K, Hester J (2022). `_desc`: Manipulate DESCRIPTION Files. R package version 1.4.2, <https://CRAN.R-project.org/package=desc>.
- **DiagrammeR 1.0.9** - Iannone R (2022). `_DiagrammeR`: Graph/Network Visualization. R package version 1.0.9, <https://CRAN.R-project.org/package=DiagrammeR>.
- **dichromat 2.0.0.1** - Lumley T (2022). `_dichromat`: Color Schemes for Dichromats. R package version 2.0-0.1, <https://CRAN.R-project.org/package=dichromat>.
- **diffobj 0.3.5** - Gaslam B (2021). `_diffobj`: Diffs for R Objects. R package version 0.3.5, <https://CRAN.R-project.org/package=diffobj>.
- **digest 0.6.29** - Lucas DEwcbA, Tuszynski J, Bengtsson H, Urbanek S, Frasca M, Lewis B, Stokely M, Muehleisen H, Murdoch D, Hester J, Wu W, Kou Q, Onkelinx T, Lang M, Simko V, Hornik K, Neal R, Bell K, de Queljo M, Suruceanu I, Denney B, Schumacher D, Chang. aW (2021). `_digest`: Create Compact Hash Digests of R Objects. R package version 0.6.29, <https://CRAN.R-project.org/package=digest>.
- **downloader 0.4** - Chang W (2015). `_downloader`: Download Files over HTTP and HTTPS. R package version 0.4, <https://CRAN.R-project.org/package=downloader>.
- **dplyr 1.0.10** - Wickham H, François R, Henry L, Müller K (2022). `_dplyr`: A Grammar of Data Manipulation. R package version 1.0.10, <https://CRAN.R-project.org/package=dplyr>.
- **DT 0.25** - Xie Y, Cheng J, Tan X (2022). `_DT`: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.25, <https://CRAN.R-project.org/package=DT>.
- **dtplyr 1.2.2** - Wickham H, Girlich M, Fairbanks M, Dickerson R (2022). `_dtplyr`: Data Table Back-End for 'dplyr'. R package version 1.2.2, <https://CRAN.R-project.org/package=dtplyr>.

- **ellipsis 0.3.2** - Wickham H (2021). `_ellipsis`: Tools for Working with ... R package version 0.3.2, <https://CRAN.R-project.org/package=ellipsis>.
- **ensemldb 2.20.2** - Rainer J, Gatto L, Weichenberger CX (2019). "ensemldb: an R package to create and use Ensembl-based annotation resources." *Bioinformatics*. doi:10.1093/bioinformatics/https://doi.org/10.1093/bioinformatics/btz031, <https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz031/5301311>.
- **evaluate 0.16** - Wickham H, Xie Y (2022). `_evaluate`: Parsing and Evaluation Tools that Provide More Details than the Default. R package version 0.16, <https://CRAN.R-project.org/package=evaluate>.
- **excelR 0.4.0** - Bista S (2020). `_excelR`: A Wrapper of the 'JavaScript' Library 'jExcel'. R package version 0.4.0, <https://CRAN.R-project.org/package=excelR>.
- **fansi 1.0.3** - Gaslam B (2022). `_fansi`: ANSI Control Sequence Aware String Functions. R package version 1.0.3, <https://CRAN.R-project.org/package=fansi>.
- **farver 2.1.1** - Pedersen T, Nicolae B, François R (2022). `_farver`: High Performance Colour Space Manipulation. R package version 2.1.1, <https://CRAN.R-project.org/package=farver>.
- **fastmap 1.1.0** - Chang W (2021). `_fastmap`: Fast Data Structures. R package version 1.1.0, <https://CRAN.R-project.org/package=fastmap>.
- **fastmatch 1.1.3** - Urbanek S (2021). `_fastmatch`: Fast 'match()' Function. R package version 1.1-3, <https://CRAN.R-project.org/package=fastmatch>.
- **fastseg 1.42.0** - Klambauer G, Schwarzbauer K, Mayr A, Mitterecker A, Clevert D, Bodenhofer U, Hochreiter S (2012). "cn.MOPS: Mixture of Poissons for Discovering Copy Number Variations in Next Generation Sequencing Data with a Low False Discovery Rate." *Nucleic Acids Research*, *40*, e69. doi:10.1093/nar/gks003 <<https://doi.org/10.1093/nar/gks003>>, <http://nar.oxfordjournals.org/content/40/9/e69.full.pdf+html>, <<http://nar.oxfordjournals.org/cont>>
- **filelock 1.0.2** - Csárdi G (2018). `_filelock`: Portable File Locking. R package version 1.0.2, <https://CRAN.R-project.org/package=filelock>.
- **flextable 0.8.1** - Gohel D, Skintzos P (2022). `_flextable`: Functions for Tabular Reporting. R package version 0.8.1, <https://CRAN.R-project.org/package=flextable>.
- **fontawesome 0.3.0** - Iannone R (2022). `_fontawesome`: Easily Work with 'Font Awesome' Icons. R package version 0.3.0, <https://CRAN.R-project.org/package=fontawesome>.
- **forcats 0.5.2** - Wickham H (2022). `_forcats`: Tools for Working with Categorical Variables (Factors). R package version 0.5.2, <https://CRAN.R-project.org/package=forcats>.
- **formatR 1.12** - Xie Y (2022). `_formatR`: Format R Code Automatically. R package version 1.12, <https://CRAN.R-project.org/package=formatR>.
- **formattable 0.2.1** - Ren K, Russell K (2021). `_formattable`: Create 'Formattable' Data Structures. R package version 0.2.1, <https://CRAN.R-project.org/package=formattable>.
- **Formula 1.2.4** - Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *Journal of Statistical Software*, *34*(1), 1-13. doi:10.18637/jss.v034.i01 <https://doi.org/10.18637/jss.v034.i01>.
- **fs 1.5.2** - Hester J, Wickham H, Csárdi G (2021). `_fs`: Cross-Platform File System Operations Based on 'libuv'. R package version 1.5.2, <https://CRAN.R-project.org/package=fs>.
- **futile.logger 1.4.3** - Rowe BLY (2016). `_futile.logger`: A Logging Utility for R. R package version 1.4.3, <https://CRAN.R-project.org/package=futile.logger>.
- **futile.options 1.0.1** - Rowe BLY (2018). `_futile.options`: Futile Options Management. R package version 1.0.1, <https://CRAN.R-project.org/package=futile.options>.
- **gargle 1.2.1** - Bryan J, Citro C, Wickham H (2022). `_gargle`: Utilities for Working with Google APIs. R package version 1.2.1, <https://CRAN.R-project.org/package=gargle>.
- **gdtools 0.2.4** - Gohel D, Wickham H, Henry L, Ooms J (2022). `_gdtools`: Utilities for Graphical Rendering. R package version 0.2.4, <https://CRAN.R-project.org/package=gdtools>.
- **generics 0.1.3** - Wickham H, Kuhn M, Vaughan D (2022). `_generics`: Common S3 Generics not Provided by Base R Methods Related to Model Fitting. R package version 0.1.3, <https://CRAN.R-project.org/package=generics>.
- **genomation 1.28.0** - Akalin A, Franke V, Vlahovicek K, Mason C, Schubeler D (2014). "genomation: a toolkit to summarize, annotate and visualize genomic intervals." *Bioinformatics*. doi:10.1093/bioinformatics/btu775 <https://doi.org/10.1093/bioinformatics/btu775>, <http://bioinformatics.oxfordjournals.org/content/early/2014/12/04/bioinformatics.btu775.long>.
- **GenomeInfoDb 1.32.2** - Arora S, Morgan M, Carlson M, Pagès H (2022). `_GenomeInfoDb`: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style. R package version 1.32.2, <https://bioconductor.org/packages/GenomeInfoDb>.
- **GenomicAlignments 1.32.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *PLoS Computational Biology*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **GenomicFeatures 1.48.3** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *PLoS Computational Biology*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **GenomicRanges 1.48.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *PLoS Computational Biology*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **ggdendro 0.1.23** - de Vries A, Ripley BD (2022). `_ggdendro`: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1.23, <https://CRAN.R-project.org/package=ggdendro>.
- **ggfortify 0.4.14** - Tang Y, Horikoshi M, Li W (2016). "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." *The R Journal*, *8*(2), 474-485. doi:10.32614/RJ-2016-060 <https://doi.org/10.32614/RJ-2016-060>, <https://doi.org/10.32614/RJ-2016-060>.
- **ggplot2 3.3.6** - Wickham H (2016). `_ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- **ggpubr 0.4.0** - Kassambara A (2020). `_ggpubr`: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0, <https://CRAN.R-project.org/package=ggpubr>.
- **ggrepel 0.9.1** - Slowikowski K (2021). `_ggrepel`: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.1, <https://CRAN.R-project.org/package=ggrepel>.

- **ggsci 2.9** - Xiao N (2018). `_ggsci`: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'. R package version 2.9, <https://CRAN.R-project.org/package=ggsci>.
- **ggsignif 0.6.3** - Constantin A, Patil I (2021). "ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'." *_PsyArxiv_*. doi:10.31234/osf.io/7awm6 <https://doi.org/10.31234/osf.io/7awm6>, <https://psyarxiv.com/7awm6>.
- **glue 1.6.2** - Hester J, Bryan J (2022). `_glue`: Interpreted String Literals. R package version 1.6.2, <https://CRAN.R-project.org/package=glue>.
- **gmp 0.6.6** - Lucas A, Scholz I, Boehme R, Jasson S, Maechler M (2022). `_gmp`: Multiple Precision Arithmetic. R package version 0.6-6, <https://CRAN.R-project.org/package=gmp>.
- **googledrive 2.0.0** - D'Agostino McGowan L, Bryan J (2021). `_googledrive`: An Interface to Google Drive. R package version 2.0.0, <https://CRAN.R-project.org/package=googledrive>.
- **googlesheets4 1.0.1** - Bryan J (2022). `_googlesheets4`: Access Google Sheets using the Sheets API V4. R package version 1.0.1, <https://CRAN.R-project.org/package=googlesheets4>.
- **gplots 3.1.3** - Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2022). `_gplots`: Various R Programming Tools for Plotting Data. R package version 3.1.3, <https://CRAN.R-project.org/package=gplots>.
- **gridBase 0.4.7** - Murrell P (2014). `_gridBase`: Integration of base and grid graphics. R package version 0.4-7, <https://CRAN.R-project.org/package=gridBase>.
- **gridExtra 2.3** - Auguie B (2017). `_gridExtra`: Miscellaneous Functions for "Grid" Graphics. R package version 2.3, <https://CRAN.R-project.org/package=gridExtra>.
- **gtable 0.3.1** - Wickham H, Pedersen T (2022). `_gtable`: Arrange 'Grobs' in Tables. R package version 0.3.1, <https://CRAN.R-project.org/package=gtable>.
- **gtools 3.9.3** - Bolker B, Warnes G, Lumley T (2022). `_gtools`: Various R Programming Tools. R package version 3.9.3, <https://CRAN.R-project.org/package=gtools>.
- **Gviz 1.40.1** - Hahne F, Ivanek R (2016). "Statistical Genomics: Methods and Protocols." In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335-351. Springer New York, New York, NY. ISBN 978-1-4939-3578-9, doi:10.1007/978-1-4939-3578-9_16 https://doi.org/10.1007/978-1-4939-3578-9_16, http://dx.doi.org/10.1007/978-1-4939-3578-9_16.
- **HardyWeinberg 1.7.5** - Graffelman J (2015). "Exploring Diallelic Genetic Markers: The HardyWeinberg Package." *_Journal of Statistical Software_*, *64*(3), 1-23. <https://www.jstatsoft.org/v64/i03/>.
- **haven 2.5.1** - Wickham H, Miller E, Smith D (2022). `_haven`: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.5.1, <https://CRAN.R-project.org/package=haven>.
- **HDF5Array 1.24.1** - Pagès H (2022). `_HDF5Array`: HDF5 backend for DelayedArray objects. R package version 1.24.1, <https://bioconductor.org/packages/HDF5Array>.
- **highr 0.9** - Xie Y, Qiu Y (2021). `_highr`: Syntax Highlighting for R Source Code. R package version 0.9, <https://CRAN.R-project.org/package=highr>.
- **Hmisc 4.7.1** - Harrell Jr F (2022). `_Hmisc`: Harrell Miscellaneous. R package version 4.7-1, <https://CRAN.R-project.org/package=Hmisc>.
- **hms 1.1.2** - Müller K (2022). `_hms`: Pretty Time of Day. R package version 1.1.2, <https://CRAN.R-project.org/package=hms>.
- **htmlTable 2.4.1** - Gordon M, Gragg S, Konings P (2022). `_htmlTable`: Advanced Tables for Markdown/HTML. R package version 2.4.1, <https://CRAN.R-project.org/package=htmlTable>.
- **htmltools 0.5.3** - Cheng J, Sievert C, Schloerke B, Chang W, Xie Y, Allen J (2022). `_htmltools`: Tools for HTML. R package version 0.5.3, <https://CRAN.R-project.org/package=htmltools>.
- **htmlwidgets 1.5.4** - Vaidyanathan R, Xie Y, Allaire J, Cheng J, Sievert C, Russell K (2021). `_htmlwidgets`: HTML Widgets for R. R package version 1.5.4, <https://CRAN.R-project.org/package=htmlwidgets>.
- **httpuv 1.6.6** - Cheng J, Chang W (2022). `_httpuv`: HTTP and WebSocket Server Library. R package version 1.6.6, <https://CRAN.R-project.org/package=httpuv>.
- **httr 1.4.4** - Wickham H (2022). `_httr`: Tools for Working with URLs and HTTP. R package version 1.4.4, <https://CRAN.R-project.org/package=httr>.
- **ids 1.0.1** - FitzJohn R (2017). `_ids`: Generate Random Identifiers. R package version 1.0.1, <https://CRAN.R-project.org/package=ids>.
- **igraph 1.3.4** - Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *_InterJournal_*, *Complex Systems*, 1695. <https://igraph.org>.
- **influenceR 0.1.0.1** - Simon J, Aditya K (2021). `_influenceR`: Software Tools to Quantify Structural Importance of Nodes in a Network. R package version 0.1.0.1, <https://CRAN.R-project.org/package=influenceR>.
- **interp 1.1.3** - Gebhardt A, Bivand R, Sinclair D (2022). `_interp`: Interpolation Methods. R package version 1.1-3, <https://CRAN.R-project.org/package=interp>.
- **IRanges 2.30.0** - Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). "Software for Computing and Annotating Genomic Ranges." *_PLoS Computational Biology_*, *9*. doi:10.1371/journal.pcbi.1003118 <https://doi.org/10.1371/journal.pcbi.1003118>, <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1003118>.
- **isoband 0.2.5** - Wilke C, Pedersen T (2021). `_isoband`: Generate Isolines and Isobands from Regularly Spaced Elevation Grids. R package version 0.2.5, <https://CRAN.R-project.org/package=isoband>.
- **jpeg 0.1.9** - Urbaneck S (2021). `_jpeg`: Read and write JPEG images. R package version 0.1-9, <https://CRAN.R-project.org/package=jpeg>.
- **jquerylib 0.1.4** - Sievert C, Cheng J (2021). `_jquerylib`: Obtain 'jQuery' as an HTML Dependency Object. R package version 0.1.4, <https://CRAN.R-project.org/package=jquerylib>.
- **jsonlite 1.8.0** - Ooms J (2014). "The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects." *_arXiv:1403.2805 [stat.CO]_*. <https://arxiv.org/abs/1403.2805>.
- **kableExtra 1.3.4** - Zhu H (2021). `_kableExtra`: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4, <https://CRAN.R-project.org/package=kableExtra>.
- **knitr 1.40** - Xie Y (2022). `_knitr`: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.40, <https://yihui.org/knitr/>.
- **labeling 0.4.2** - Talbot, J (2020). `_labeling`: Axis Labeling. R package version 0.4.2, <https://CRAN.R-project.org/package=labeling>.
- **lambda.r 1.2.4** - Rowe BLY (2019). `_lambda.r`: Modeling Data with Functional Programming.

- R package version 1.2.4, <https://CRAN.R-project.org/package=lambda.r>.
- **later 1.3.0** - Chang W, Cheng J (2021). `_later`: Utilities for Scheduling Functions to Execute Later with Event Loops. R package version 1.3.0, <https://CRAN.R-project.org/package=later>.
 - **latticeExtra 0.6.30** - Sarkar D, Andrews F (2022). `_latticeExtra`: Extra Graphical Utilities Based on Lattice. R package version 0.6-30, <https://CRAN.R-project.org/package=latticeExtra>.
 - **lazyeval 0.2.2** - Wickham H (2019). `_lazyeval`: Lazy (Non-Standard) Evaluation. R package version 0.2.2, <https://CRAN.R-project.org/package=lazyeval>.
 - **lifecycle 1.0.2** - Henry L, Wickham H (2022). `_lifecycle`: Manage the Life Cycle of your Package Functions. R package version 1.0.2, <https://CRAN.R-project.org/package=lifecycle>.
 - **limma 3.52.2** - Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." *Nucleic Acids Research*, *43*(7), e47. doi:10.1093/nar/gkv007 <https://doi.org/10.1093/nar/gkv007>.
 - **lme4 1.1.30** - Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01 <https://doi.org/10.18637/jss.v067.i01>.
 - **locfit 1.5.9.6** - Loader C (2022). `_locfit`: Local Regression, Likelihood and Density Estimation. R package version 1.5-9.6, <https://CRAN.R-project.org/package=locfit>.
 - **logger 0.2.2** - Daróczy G (2021). `_logger`: A Lightweight, Modern and Flexible Logging Utility. R package version 0.2.2, <https://CRAN.R-project.org/package=logger>.
 - **lubridate 1.8.0** - Grolemund G, Wickham H (2011). "Dates and Times Made Easy with lubridate." *Journal of Statistical Software*, *40*(3), 1-25. <https://www.jstatsoft.org/v40/i03/>.
 - **magrittr 2.0.3** - Bache S, Wickham H (2022). `_magrittr`: A Forward-Pipe Operator for R. R package version 2.0.3, <https://CRAN.R-project.org/package=magrittr>.
 - **maptools 1.1.4** - Bivand R, Lewin-Koh N (2022). `_maptools`: Tools for Handling Spatial Objects. R package version 1.1-4, <https://CRAN.R-project.org/package=maptools>.
 - **MASS 7.3.58.1** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
 - **Matrix 1.5.1** - Bates D, Maechler M, Jagan M (2022). `_Matrix`: Sparse and Dense Matrix Classes and Methods. R package version 1.5-1, <https://CRAN.R-project.org/package=Matrix>.
 - **MatrixGenerics 1.8.1** - Ahlmann-Eltze C, Hickey P, Pagès H (2022). `_MatrixGenerics`: S4 Generic Summary Statistic Functions that Operate on Matrix-Like Objects. R package version 1.8.1, <https://bioconductor.org/packages/MatrixGenerics>.
 - **MatrixModels 0.5.1** - Bates D, Maechler M (2022). `_MatrixModels`: Modelling with Sparse and Dense Matrices. R package version 0.5-1, <https://CRAN.R-project.org/package=MatrixModels>.
 - **matrixStats 0.62.0** - Bengtsson H (2022). `_matrixStats`: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.62.0, <https://CRAN.R-project.org/package=matrixStats>.
 - **mclust 5.4.10** - Scrucca L, Fop M, Murphy TB, Raftery AE (2016). "mclust 5: clustering,

- classification and density estimation using Gaussian finite mixture models." *The R Journal*, *8*(1), 289-317. <https://doi.org/10.32614/RJ-2016-021>.
- **memoise 2.0.1** - Wickham H, Hester J, Chang W, Müller K, Cook D (2021). `_memoise`: 'Memoisation' of Functions. R package version 2.0.1, <https://CRAN.R-project.org/package=memoise>.
 - **mice 3.14.0** - van Buuren S, Groothuis-Oudshoorn K (2011). "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software*, *45*(3), 1-67. doi:10.18637/jss.v045.i03 <https://doi.org/10.18637/jss.v045.i03>.
 - **mime 0.12** - Xie Y (2021). `_mime`: Map Filenames to MIME Types. R package version 0.12, <https://CRAN.R-project.org/package=mime>.
 - **minqa 1.2.4** - Bates D, Mullen KM, Nash JC, Varadhan R (2014). `_minqa`: Derivative-free optimization algorithms by quadratic approximation. R package version 1.2.4, <https://CRAN.R-project.org/package=minqa>.
 - **mnormt 2.1.0** - Azzalini A, Genz A (2022). `_The R package mnormt`: The multivariate normal and t distributions (version 2.1.0). <http://azzalini.stat.unipd.it/SW/Pkg-mnormt/>.
 - **modelr 0.1.9** - Wickham H (2022). `_modelr`: Modelling Functions that Work with the Pipe. R package version 0.1.9, <https://CRAN.R-project.org/package=modelr>.
 - **munsell 0.5.0** - Wickham C (2018). `_munsell`: Utilities for Using Munsell Colours. R package version 0.5.0, <https://CRAN.R-project.org/package=munsell>.
 - **mvtnorm 1.1.3** - Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021). `_mvtnorm`: Multivariate Normal and t Distributions. R package version 1.1-3, <https://CRAN.R-project.org/package=mvtnorm>.
 - **nlme 3.1.159** - Pinheiro J, Bates D, R Core Team (2022). `_nlme`: Linear and Nonlinear Mixed Effects Models. R package version 3.1-159, <https://CRAN.R-project.org/package=nlme>.
 - **numDeriv 2016.8.1.1** - Gilbert P, Varadhan R (2019). `_numDeriv`: Accurate Numerical Derivatives. R package version 2016.8-1.1, <https://CRAN.R-project.org/package=numDeriv>.
 - **officer 0.4.4** - Gohel D (2022). `_officer`: Manipulation of Microsoft Word and PowerPoint Documents. R package version 0.4.4, <https://CRAN.R-project.org/package=officer>.
 - **openssl 2.0.3** - Ooms J (2022). `_openssl`: Toolkit for Encryption, Signatures and Certificates Based on OpenSSL. R package version 2.0.3, <https://CRAN.R-project.org/package=openssl>.
 - **openxlsx 4.2.5** - Schaubberger P, Walker A (2021). `_openxlsx`: Read, Write and Edit xlsx Files. R package version 4.2.5, <https://CRAN.R-project.org/package=openxlsx>.
 - **pbkrtest 0.5.1** - Halekoh U, Højsgaard S (2014). "A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest." *Journal of Statistical Software*, *59*(9), 1-30. <https://www.jstatsoft.org/v59/i09/>.
 - **pdftools 3.3.0** - Ooms J (2022). `_pdftools`: Text Extraction, Rendering and Converting of PDF Documents. R package version 3.3.0, <https://CRAN.R-project.org/package=pdfutils>.
 - **permut 0.9.7** - Simpson G (2022). `_permut`: Functions for Generating Restricted Permutations of Data. R package version 0.9-7, <https://CRAN.R-project.org/package=permut>.
 - **phangorn 2.10.0** - Schliep K (2011). "phangorn: phylogenetic analysis in R." *Bioinformatics*, *27*(4), 592-593. doi:10.1093/bioinformatics/btq706 <https://doi.org/10.1093/bioinformatics/btq706>.

- **pillar 1.8.1** - Müller K, Wickham H (2022). `_pillar`: Coloured Formatting for Columns_. R package version 1.8.1, <https://CRAN.R-project.org/package=pillar>.
- **pixmap 0.4.12** - Bivand R, Leisch F, Maechler M (2021). `_pixmap`: Bitmap Images / Pixel Maps_. R package version 0.4-12, <https://CRAN.R-project.org/package=pixmap>.
- **pkgconfig 2.0.3** - Csárdi G (2019). `_pkgconfig`: Private Configuration for 'R' Packages_. R package version 2.0.3, <https://CRAN.R-project.org/package=pkgconfig>.
- **pkgload 1.3.0** - Wickham H, Chang W, Hester J, Henry L (2022). `_pkgload`: Simulate Package Installation and Attach_. R package version 1.3.0, <https://CRAN.R-project.org/package=pkgload>.
- **plogr 0.2.0** - Müller K (2018). `_plogr`: The 'plogr' C++ Logging Library_. R package version 0.2.0, <https://CRAN.R-project.org/package=plogr>.
- **plotly 4.10.0** - Sievert C (2020). `_Interactive Web-Based Data Visualization with R, plotly, and shiny_`. Chapman and Hall/CRC. ISBN 9781138331457, <https://plotly-r.com>.
- **plyr 1.8.7** - Wickham H (2011). "The Split-Apply-Combine Strategy for Data Analysis." `_Journal of Statistical Software_`, *40*(1), 1-29. <https://www.jstatsoft.org/v40/i01/>.
- **png 0.1.7** - Urbanek S (2013). `_png`: Read and write PNG images_. R package version 0.1-7, <https://CRAN.R-project.org/package=png>.
- **polynom 1.4.1** - Venables B, Hornik K, Maechler M (2022). `_polynom`: A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations_. R package version 1.4-1. S original by Bill Venables, packages for R by Kurt Hornik and Martin Maechler, <https://CRAN.R-project.org/package=polynom>.
- **praise 1.0.0** - Csardi G, Sorhus S (2015). `_praise`: Praise Users_. R package version 1.0.0, <https://CRAN.R-project.org/package=praise>.
- **prettyunits 1.1.1** - Csardi G (2020). `_prettyunits`: Pretty, Human Readable Formatting of Quantities_. R package version 1.1.1, <https://CRAN.R-project.org/package=prettyunits>.
- **processx 3.7.0** - Csárdi G, Chang W (2022). `_processx`: Execute and Control System Processes_. R package version 3.7.0, <https://CRAN.R-project.org/package=processx>.
- **progress 1.2.2** - Csárdi G, FitzJohn R (2019). `_progress`: Terminal Progress Bars_. R package version 1.2.2, <https://CRAN.R-project.org/package=progress>.
- **promises 1.2.0.1** - Cheng J (2021). `_promises`: Abstractions for Promise-Based Asynchronous Programming_. R package version 1.2.0.1, <https://CRAN.R-project.org/package=promises>.
- **ProtGenerics 1.28.0** - Gatto L, Rainer J (2022). `_ProtGenerics`: Generic infrastructure for Bioconductor mass spectrometry packages_. R package version 1.28.0, <https://github.com/RforMassSpectrometry/ProtGenerics>.
- **ps 1.7.1** - Loden J, Daeschler D, Rodola' G, Csárdi G (2022). `_ps`: List, Query, Manipulate System Processes_. R package version 1.7.1, <https://CRAN.R-project.org/package=ps>.
- **psych 2.2.5** - Revelle W (2022). `_psych`: Procedures for Psychological, Psychometric, and Personality Research_. Northwestern University, Evanston, Illinois. R package version 2.2.5, <https://CRAN.R-project.org/package=psych>.
- **purrr 0.3.4** - Henry L, Wickham H (2020). `_purrr`: Functional Programming Tools_. R package version 0.3.4, <https://CRAN.R-project.org/package=purrr>.
- **qpdf 1.2.0** - Ooms J (2022). `_qpdf`: Split, Combine and Compress PDF Files_. R package version 1.2.0, <https://CRAN.R-project.org/package=qpdf>.
- **quadprog 1.5.8** - dpodi/LINPACK) SobBAtRpBAW<FcfCM (2019). `_quadprog`: Functions to Solve Quadratic Programming Problems_. R package version 1.5-8, <<https://CRAN.R-project.org/package=quadprog>>.
- **quantreg 5.94** - Koenker R (2022). `_quantreg`: Quantile Regression_. R package version 5.94, <https://CRAN.R-project.org/package=quantreg>.
- **qvalue 2.28.0** - Storey JD, Bass AJ, Dabney A, Robinson D (2022). `_qvalue`: Q-value estimation for false discovery rate control_. R package version 2.28.0, <http://github.com/jdstorey/qvalue>.
- **R.methodsS3 1.8.2** - Bengtsson H (2003). "The R.oo package - Object-Oriented Programming with References Using Standard R Code." In Hornik K, Leisch F, Zeileis A (eds.), `_Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)_`. <https://www.r-project.org/conferences/DSC-2003/Proceedings/Bengtsson.pdf>.
- **R.oo 1.25.0** - Bengtsson H (2003). "The R.oo package - Object-Oriented Programming with References Using Standard R Code." In Hornik K, Leisch F, Zeileis A (eds.), `_Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)_`. <https://www.r-project.org/conferences/DSC-2003/Proceedings/Bengtsson.pdf>.
- **R.utils 2.12.0** - Bengtsson H (2022). `_R.utils`: Various Programming Utilities_. R package version 2.12.0, <https://CRAN.R-project.org/package=R.utils>.
- **R6 2.5.1** - Chang W (2021). `_R6`: Encapsulated Classes with Reference Semantics_. R package version 2.5.1, <https://CRAN.R-project.org/package=R6>.
- **rappdirs 0.3.3** - Ratnakumar S, Mick T, Davis T (2021). `_rappdirs`: Application Directories: Determine Where to Save Data, Caches, and Logs_. R package version 0.3.3, <https://CRAN.R-project.org/package=rappdirs>.
- **RBGL 1.72.0** - Carey V, Long L, Gentleman R (2022). `_RBGL`: An interface to the BOOST graph library_. R package version 1.72.0, <http://www.bioconductor.org>.
- **RColorBrewer 1.1.3** - Neuwirth E (2022). `_RColorBrewer`: ColorBrewer Palettes_. R package version 1.1-3, <https://CRAN.R-project.org/package=RColorBrewer>.
- **Rcpp 1.0.9** - Eddelbuettel D, François R (2011). "Rcpp: Seamless R and C++ Integration." `_Journal of Statistical Software_`, *40*(8), 1-18. doi:10.18637/jss.v040.i08 <https://doi.org/10.18637/jss.v040.i08>.
- **RcppEigen 0.3.3.9.2** - Bates D, Eddelbuettel D (2013). "Fast and Elegant Numerical Linear Algebra Using the RcppEigen Package." `_Journal of Statistical Software_`, *52*(5), 1-24. doi:10.18637/jss.v052.i05 <https://doi.org/10.18637/jss.v052.i05>.
- **RCurl 1.98.1.8** - Temple Lang D (2022). `_RCurl`: General Network (HTTP/FTP/...) Client Interface for R_. R package version 1.98-1.8, <https://CRAN.R-project.org/package=RCurl>.
- **readr 2.1.2** - Wickham H, Hester J, Bryan J (2022). `_readr`: Read Rectangular Text Data_. R package version 2.1.2, <https://CRAN.R-project.org/package=readr>.
- **readxl 1.4.1** - Wickham H, Bryan J (2022). `_readxl`: Read Excel Files_. R package version 1.4.1, <https://CRAN.R-project.org/package=readxl>.
- **regioneR 1.28.0** - Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R (2016). "regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests." `_Bioinformatics_`, *32*(2), 289-291. doi:10.1093/bioinformatics/https://doi.org/10.1093/bioinformatics/btv562
- **rematch 1.0.1** - Csardi G (2016). `_rematch`: Match Regular Expressions with a Nicer 'API'_. R package version 1.0.1, <https://CRAN.R-project.org/package=rematch>.

- **rematch2 2.1.2** - Csárdi G (2020). `_rematch2`: Tidy Output from Regular Expression Matching_. R package version 2.1.2, <https://CRAN.R-project.org/package=rematch2>.
- **renv 0.15.5** - Ushey K (2022). `_renv`: Project Environments_. R package version 0.15.5, <https://CRAN.R-project.org/package=renv>.
- **reprex 2.0.2** - Bryan J, Hester J, Robinson D, Wickham H, Dervieux C (2022). `_reprex`: Prepare Reproducible Example Code via the Clipboard_. R package version 2.0.2, <https://CRAN.R-project.org/package=reprex>.
- **reshape2 1.4.4** - Wickham H (2007). "Reshaping Data with the reshape Package." *Journal of Statistical Software*_. *21*(12), 1-20. <http://www.jstatsoft.org/v21/i12/>.
- **restfulr 0.0.15** - Lawrence M (2022). `_restfulr`: R Interface to RESTful Web Services_. R package version 0.0.15, <https://CRAN.R-project.org/package=restfulr>.
- **rhdf5 2.40.0** - Fischer B, Smith M, Pau G (2022). `_rhdf5`: R Interface to HDF5_. R package version 2.40.0, <https://github.com/grimbough/rhdf5>.
- **rhdf5filters 1.8.0** - Smith M (2022). `_rhdf5filters`: HDF5 Compression Filters_. R package version 1.8.0, <https://github.com/grimbough/rhdf5filters>.
- **Rhdf5lib 1.18.2** - Smith M (2022). `_Rhdf5lib`: hdf5 library as an R package_. R package version 1.18.2, <https://github.com/grimbough/Rhdf5lib>.
- **rjson 0.2.21** - Couture-Beil A (2022). `_rjson` for R_. R package version 0.2.21, <https://CRAN.R-project.org/package=rjson>.
- **rlang 1.0.5** - Henry L, Wickham H (2022). `_rlang`: Functions for Base Types and Core R and 'Tidyverse' Features_. R package version 1.0.5, <https://CRAN.R-project.org/package=rlang>.
- **rlist 0.4.6.2** - Ren K (2021). `_rlist`: A Toolbox for Non-Tabular Data Manipulation_. R package version 0.4.6.2, <https://CRAN.R-project.org/package=rlist>.
- **rmarkdown 2.16** - Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2022). `_rmarkdown`: Dynamic Documents for R_. R package version 2.16, <https://github.com/rstudio/rmarkdown>.
- **Rmisc 1.5.1** - Hope RM (2022). `_Rmisc`: Ryan Miscellaneous_. R package version 1.5.1, <https://CRAN.R-project.org/package=Rmisc>.
- **rprojroot 2.0.3** - Müller K (2022). `_rprojroot`: Finding Files in Project Subdirectories_. R package version 2.0.3, <https://CRAN.R-project.org/package=rprojroot>.
- **Rsamtools 2.12.0** - Morgan M, Pagès H, Obenchain V, Hayden N (2022). `_Rsamtools`: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import_. R package version 2.12.0, <https://bioconductor.org/packages/Rsamtools>.
- **RSQLite 2.2.17** - Müller K, Wickham H, James DA, Falcon S (2022). `_RSQLite`: SQLite Interface for R_. R package version 2.2.17, <https://CRAN.R-project.org/package=RSQLite>.
- **rstatix 0.7.0** - Kassambara A (2021). `_rstatix`: Pipe-Friendly Framework for Basic Statistical Tests_. R package version 0.7.0, <https://CRAN.R-project.org/package=rstatix>.
- **rstudioapi 0.14** - Ushey K, Allaire J, Wickham H, Ritchie G (2022). `_rstudioapi`: Safely Access the RStudio API_. R package version 0.14, <https://CRAN.R-project.org/package=rstudioapi>.
- **rtracklayer 1.56.1** - Lawrence M, Gentleman R, Carey V (2009). "rtracklayer: an R package for interfacing with genome browsers." *Bioinformatics*_. *25*, 1841-1842. doi:10.1093/bioinformatics. <https://doi.org/10.1093/bioinformatics/btp328>, <http://bioinformatics.oxfordjournals.org/content/25/14/1841.abstract>.

- **rvest 1.0.3** - Wickham H (2022). `_rvest`: Easily Harvest (Scrape) Web Pages_. R package version 1.0.3, <https://CRAN.R-project.org/package=rvest>.
- **S4Vectors 0.34.0** - Pagès H, Lawrence M, Aboyou P (2022). `_S4Vectors`: Foundation of vector-like and list-like containers in Bioconductor_. R package version 0.34.0, <https://bioconductor.org/packages/S4Vectors>.
- **sangeranalyseR 1.6.1** - Chao K, Barton K, Palmer S, Lanfear R (2021). "sangeranalyseR: simple and interactive analysis of Sanger sequencing data in R." *Genome Biology and Evolution*_. doi:10.1093/gbe/evab028 <https://doi.org/10.1093/gbe/evab028>.
- **sangerseqR 1.32.0** - Hill JT, Demarest BL, Bisgrove BW, Su Y, Smith M, Yost HJ (2014). "Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products." *Developmental Dynamics*_. doi:10.1002/dvdy.24183. <https://doi.org/10.1002/dvdy.24183>..
- **sass 0.4.2** - Cheng J, Mastny T, Iannone R, Schloerke B, Sievert C (2022). `_sass`: Syntactically Awesome Style Sheets ('Sass')_. R package version 0.4.2, <https://CRAN.R-project.org/package=sass>.
- **scales 1.2.1** - Wickham H, Seidel D (2022). `_scales`: Scale Functions for Visualization_. R package version 1.2.1, <https://CRAN.R-project.org/package=scales>.
- **selectr 0.4.2** - Potter S (2012). "Introducing the selectr Package." The University of Auckland, Auckland, New Zealand. <http://stattech.wordpress.fos.auckland.ac.nz/2012-10-introducing-the-selectr-package/>.
- **shiny 1.7.2** - Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2022). `_shiny`: Web Application Framework for R_. R package version 1.7.2, <https://CRAN.R-project.org/package=shiny>.
- **shinycssloaders 1.0.0** - Sali A, Attali D (2020). `_shinycssloaders`: Add Loading Animations to a 'shiny' Output While It's Recalculating_. R package version 1.0.0, <https://CRAN.R-project.org/package=shinycssloaders>.
- **shinydashboard 0.7.2** - Chang W, Borges Ribeiro B (2021). `_shinydashboard`: Create Dashboards with 'Shiny'_. R package version 0.7.2, <https://CRAN.R-project.org/package=shinydashboard>.
- **shinyjs 2.1.0** - Attali D (2021). `_shinyjs`: Easily Improve the User Experience of Your Shiny Apps in Seconds_. R package version 2.1.0, <https://CRAN.R-project.org/package=shinyjs>.
- **shinythemes 1.2.0** - Chang W (2021). `_shinythemes`: Themes for Shiny_. R package version 1.2.0, <https://CRAN.R-project.org/package=shinythemes>.
- **shinyWidgets 0.7.3** - Perrier V, Meyer F, Granjon D (2022). `_shinyWidgets`: Custom Inputs Widgets for Shiny_. R package version 0.7.3, <https://CRAN.R-project.org/package=shinyWidgets>.
- **snow 0.4.4** - Tierney L, Rossini AJ, Li N, Sevcikova H (2021). `_snow`: Simple Network of Workstations_. R package version 0.4-4, <https://CRAN.R-project.org/package=snow>.
- **sourcetools 0.1.7** - Ushey K (2018). `_sourcetools`: Tools for Reading, Tokenizing and Parsing R Code_. R package version 0.1.7, <https://CRAN.R-project.org/package=sourcetools>.
- **sp 1.5.0** - Pebesma EJ, Bivand RS (2005). "Classes and methods for spatial data in R." *R News*_. *5*(2), 9-13. <https://CRAN.R-project.org/doc/Rnews/>.
- **SparseM 1.81** - Koenker R (2021). `_SparseM`: Sparse Linear Algebra_. R package version

1.81, <https://CRAN.R-project.org/package=SparseM>.

- **sparseMatrixStats 1.8.0** - Ahlmann-Eltze C (2022). `_sparseMatrixStats`: Summary Statistics for Rows and Columns of Sparse Matrices_. R package version 1.8.0, <https://github.com/const-ae/sparseMatrixStats>.
- **stringi 1.7.8** - Gagolewski M (2022). "stringi: Fast and portable character string processing in R." *Journal of Statistical Software*_, *103*(2), 1-59. doi:10.18637/jss.v103.i02 <https://doi.org/10.18637/jss.v103.i02>.
- **stringr 1.4.1** - Wickham H (2022). `_stringr`: Simple, Consistent Wrappers for Common String Operations_. R package version 1.4.1, <https://CRAN.R-project.org/package=stringr>.
- **SummarizedExperiment 1.26.1** - Morgan M, Obenchain V, Hester J, Pagès H (2022). `_SummarizedExperiment`: SummarizedExperiment container_. R package version 1.26.1, <https://bioconductor.org/packages/SummarizedExperiment>.
- **survival 3.4.0** - Therneau T (2022). `_A` Package for Survival Analysis in R_. R package version 3.4.0, <https://CRAN.R-project.org/package=survival>.
- **svglite 2.1.0** - Wickham H, Henry L, Pedersen T, Luciani T, Decorde M, Lise V (2022). `_svglite`: An 'SVG' Graphics Device_. R package version 2.1.0, <https://CRAN.R-project.org/package=svglite>.
- **sys 3.4** - Ooms J (2020). `_sys`: Powerful and Reliable Tools for Running System Commands in R_. R package version 3.4, <https://CRAN.R-project.org/package=sys>.
- **systemfonts 1.0.4** - Pedersen T, Ooms J, Govett D (2022). `_systemfonts`: System Native Font Finding_. R package version 1.0.4, <https://CRAN.R-project.org/package=systemfonts>.
- **testthat 3.1.4** - Wickham H (2011). "testthat: Get Started with Testing." *The R Journal*_, *3*, 5-10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.
- **tibble 3.1.8** - Müller K, Wickham H (2022). `_tibble`: Simple Data Frames_. R package version 3.1.8, <https://CRAN.R-project.org/package=tibble>.
- **tidyr 1.2.1** - Wickham H, Girlich M (2022). `_tidyr`: Tidy Messy Data_. R package version 1.2.1, <https://CRAN.R-project.org/package=tidyr>.
- **tidyselect 1.1.2** - Henry L, Wickham H (2022). `_tidyselect`: Select from a Set of Strings_. R package version 1.1.2, <https://CRAN.R-project.org/package=tidyselect>.
- **tidyverse 1.3.2** - Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
- **tinytex 0.41** - Xie Y (2022). `_tinytex`: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents_. R package version 0.41, <https://github.com/rstudio/tinytex>.
- **truncnorm 1.0.8** - Mersmann O, Trautmann H, Steuer D, Bornkamp B (2018). `_truncnorm`: Truncated Normal Distribution_. R package version 1.0-8, <https://CRAN.R-project.org/package=truncnorm>.
- **tzdb 0.3.0** - Vaughan D (2022). `_tzdb`: Time Zone Database Information_. R package version 0.3.0, <https://CRAN.R-project.org/package=tzdb>.
- **utf8 1.2.2** - Perry PO (2021). `_utf8`: Unicode Text Processing_. R package version 1.2.2,

<https://CRAN.R-project.org/package=utf8>.

- **uuid 1.1.0** - Urbaneck S, Ts'o T (2022). `_uuid`: Tools for Generating and Handling of UUIDs_. R package version 1.1-0, <https://CRAN.R-project.org/package=uuid>.
- **VariantAnnotation 1.42.1** - Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M (2014). "VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants." *Bioinformatics*_, *30*(14), 2076-2078. doi:10.1093/bioinformatics/btu168 <https://doi.org/10.1093/bioinformatics/btu168>.
- **vctrs 0.4.1** - Wickham H, Henry L, Vaughan D (2022). `_vctrs`: Vector Helpers_. R package version 0.4.1, <https://CRAN.R-project.org/package=vctrs>.
- **VennDiagram 1.7.3** - Chen H (2022). `_VennDiagram`: Generate High-Resolution Venn and Euler Plots_. R package version 1.7.3, <https://CRAN.R-project.org/package=VennDiagram>.
- **viridis 0.6.2** - Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric (2021). `_viridis` - Colorblind-Friendly Color Maps for R_. doi:10.5281/zenodo.467942 <https://doi.org/10.5281/zenodo.467942>, <https://rpubs.io/467942>, <https://sjmga.github.io/viridis/>.
- **viridisLite 0.4.1** - Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric (2022). `_viridis` - Colorblind-Friendly Color Maps for R_. doi:10.5281/zenodo.4679424, <https://doi.org/10.5281/zenodo.4679424>, <https://rpubs.io/4679424>, <https://sjmga.github.io/viridis/>.
- **visNetwork 2.1.0** - Almende B.V. and Contributors, Thieurmel B, Robert T (2021). `_visNetwork`: Network Visualization using 'vis.js' Library_. R package version 2.1.0, <https://CRAN.R-project.org/package=visNetwork>.
- **vroom 1.5.7** - Hester J, Wickham H, Bryan J (2021). `_vroom`: Read and Write Rectangular Text Data Quickly_. R package version 1.5.7, <https://CRAN.R-project.org/package=vroom>.
- **waldo 0.4.0** - Wickham H (2022). `_waldo`: Find Differences Between R Objects_. R package version 0.4.0, <https://CRAN.R-project.org/package=waldo>.
- **webshot 0.5.3** - Chang W (2022). `_webshot`: Take Screenshots of Web Pages_. R package version 0.5.3, <https://CRAN.R-project.org/package=webshot>.
- **withr 2.5.0** - Hester J, Henry L, Müller K, Ushey K, Wickham H, Chang W (2022). `_withr`: Run Code 'With' Temporarily Modified Global State_. R package version 2.5.0, <https://CRAN.R-project.org/package=withr>.
- **writexl 1.4.0** - Ooms J (2021). `_writexl`: Export Data Frames to Excel 'xlsx' Format_. R package version 1.4.0, <https://CRAN.R-project.org/package=writexl>.
- **xfun 0.33** - Xie Y (2022). `_xfun`: Supporting Functions for Packages Maintained by 'Yihui Xie'_. R package version 0.33, <https://CRAN.R-project.org/package=xfun>.
- **XML 3.99.0.10** - Temple Lang D (2022). `_XML`: Tools for Parsing and Generating XML Within R and S-Plus_. R package version 3.99-0.10, <https://CRAN.R-project.org/package=XML>.
- **xml2 1.3.3** - Wickham H, Hester J, Ooms J (2021). `_xml2`: Parse XML_. R package version 1.3.3, <https://CRAN.R-project.org/package=xml2>.
- **xtable 1.8.4** - Dahl D, Scott D, Roosen C, Magnusson A, Swinton J (2019). `_xtable`: Export Tables to LaTeX or HTML_. R package version 1.8-4, <https://CRAN.R-project.org/package=xtable>.
- **XVector 0.36.0** - Pagès H, Aboyoun P (2022). `_XVector`: Foundation of external vector representation and manipulation in Bioconductor_. R package version 0.36.0, <https://>

bioconductor.org/packages/XVector.

- **yaml 2.3.5** - Garbett SP, Stephens J, Simonov K, Xie Y, Dong Z, Wickham H, Horner J, reikoch, Beasley W, O'Connor B, Warnes GR, Quinn M, Kamvar ZN (2022). `_yaml`: Methods to Convert R Data to YAML and Back. R package version 2.3.5, <https://CRAN.R-project.org/package=yaml>.
- **zeallot 0.1.0** - Teeter N (2018). `_zeallot`: Multiple, Unpacking, and Deconstructing Assignment. R package version 0.1.0, <https://CRAN.R-project.org/package=zeallot>.
- **zip 2.2.1** - Csárdi G, Podgórski K, Geldreich R (2022). `_zip`: Cross-Platform 'zip' Compression. R package version 2.2.1, <https://CRAN.R-project.org/package=zip>.
- **zlibbioc 1.42.0** - Morgan M (2022). `_zlibbioc`: An R packaged zlib-1.2.5. R package version 1.42.0, <https://bioconductor.org/packages/zlibbioc>.
- **base 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **class 7.3.20** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **cluster 2.1.4** - Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2022). `_cluster`: Cluster Analysis Basics and Extensions. R package version 2.1.4 - For new features, see the 'Changelog' file (in the package source), <https://CRAN.R-project.org/package=cluster>.
- **codetools 0.2.18** - Tierney L (2020). `_codetools`: Code Analysis Tools for R. R package version 0.2-18, <https://CRAN.R-project.org/package=codetools>.
- **compiler 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **datasets 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **foreign 0.8.82** - R Core Team (2022). `_foreign`: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', R package version 0.8-82, <https://CRAN.R-project.org/package=foreign>.
- **graphics 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **grDevices 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **grid 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **KernSmooth 2.23.20** - Wand M (2021). `_KernSmooth`: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package version 2.23-20, <https://CRAN.R-project.org/package=KernSmooth>.
- **lattice 0.20.45** - Sarkar D (2008). `_Lattice`: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5, <http://lmdvr.r-forge.r-project.org>.
- **MASS 7.3.58.1** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/>

[MASS4/](https://www.stats.ox.ac.uk/pub/MASS4/).

- **Matrix 1.5.1** - Bates D, Maechler M, Jagan M (2022). `_Matrix`: Sparse and Dense Matrix Classes and Methods. R package version 1.5-1, <https://CRAN.R-project.org/package=Matrix>.
- **methods 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **nlme 3.1.159** - Pinheiro J, Bates D, R Core Team (2022). `_nlme`: Linear and Nonlinear Mixed Effects Models. R package version 3.1-159, <https://CRAN.R-project.org/package=nlme>.
- **nnet 7.3.17** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **parallel 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **rpart 4.1.16** - Therneau T, Atkinson B (2022). `_rpart`: Recursive Partitioning and Regression Trees. R package version 4.1.16, <https://CRAN.R-project.org/package=rpart>.
- **spatial 7.3.15** - Venables WN, Ripley BD (2002). `_Modern Applied Statistics with S`, Fourth edition. Springer, New York. ISBN 0-387-95457-0, <https://www.stats.ox.ac.uk/pub/MASS4/>.
- **splines 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **stats 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **stats4 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **survival 3.4.0** - Therneau T (2022). `_A Package for Survival Analysis in R`. R package version 3.4-0, <https://CRAN.R-project.org/package=survival>.
- **tcltk 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **tools 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- **utils 4.2.1** - R Core Team (2022). `_R`: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

STUDY OF DNA METHYLATION MODIFICATIONS: FROM DYNAMICS DURING THE DEDIFFERENTIATION INTO BREAST CANCER STEM CELLS, TO THE DEVELOPMENT OF THE R-BASED TOOL ABSP, ANALYSIS OF BISULFITE SEQUENCING PCR

Cancer stem cells (CSCs) form a tumoral subpopulation characterized by self-renewal abilities, pluripotency, therapeutic resistance mechanisms, and tumor initiation capacities, and are therefore a major cause of cancer recurrence after treatments. Moreover, the non-cancer stem cells (non-CSCs) are able to dedifferentiate into CSCs, in response to stress, especially to antitumor treatments such as radiotherapy, thus reinforcing the therapeutic resistance of cancer. In addition, epigenetic marks such as DNA methylation are known to contribute to the regulation of stemness properties and could be involved in the reacquisition of a CSC phenotype.

To evaluate DNA methylation modifications occurring throughout the radio-induced dedifferentiation of non-CSCs into CSCs in the breast cancer model, a Reduced Representation Bisulfite Sequencing (RRBS) analysis of the different tumor subpopulations was carried out. The analysis of RRBS data led to the identification of over 2,000 Differentially Methylated Regions (DMRs) undergoing methylation changes from non-CSC to radio-induced CSC. Among them, 35 present a methylation profile across the populations consistent with a potential contribution to radio-induced dedifferentiation. Five regions, associated with the FSCN1, CHRNA6, CDH7, CD9, and PRKAR1B genes, were selected for further validation. Genes regulated by these methylation changes could serve as new therapeutic targets to specifically inhibit the non-CSC to CSC phenotypic switch and prevent the enrichment in CSCs, reducing the risk of cancer relapse.

To validate identified methylation differences, the Bisulfite Sequencing PCR (BSP) method was chosen as it is the most convenient and accessible technique to quantify locus-specific methylation levels. Due to a lack of efficient tools to analyze BSP results from both approaches (direct-BSP and cloning-BSP), the ABSP R-based tool, standing for Analysis of Bisulfite Sequencing PCR, was developed. This tool provides a complete, automated, and user-friendly workflow to compute methylation percentages and compare methylation differences between samples. ABSP is available for download, along with associated data, at <https://github.com/ABSP-methylation-tool/ABSP>. Altogether, this work highlights the importance of DNA methylation within CSC plasticity and the room for tools to improve its analysis.

KEYWORDS: *breast cancer, radiotherapy, cancer stem cells, DNA methylation, bisulfite sequencing, R language*

ÉTUDE DES MODIFICATIONS DE MÉTHYLATION DE L'ADN : DES DYNAMIQUES AU COURS DE LA DÉDIFFÉRENCIATION EN CELLULES SOUCHES CANCÉREUSES DE SEIN, AU DÉVELOPPEMENT DE L'OUTIL ABSP, *Analysis of Bisulfite Sequencing PCR*, SOUS R

Les Cellules Souches Cancéreuses (CSC) forment une sous-population tumorale caractérisée par des capacités d'auto-renouvellement, de pluripotence, d'initiation tumorale et présentent une résistance thérapeutique accrue. Elles sont donc une cause majeure de récurrence du cancer. De plus, les cellules cancéreuses non-souches sont capables de se différencier en CSC en réponse à un stress, notamment aux traitements anticancéreux comme la radiothérapie, renforçant ainsi la résistance thérapeutique de la tumeur. Nous avons fait l'hypothèse que les marques épigénétiques telles que la méthylation de l'ADN, connues comme contribuant à la régulation des propriétés souches, seraient impliquées dans la réacquisition d'un phénotype CSC.

Afin d'évaluer les modifications de méthylation de l'ADN au cours de la différenciation radio-induite des cellules non-CSC en CSC dans le modèle de cancer du sein, une analyse de *Reduced Representation Bisulfite Sequencing* (RRBS) des différentes sous-populations tumorales a été réalisée. Cette analyse a permis d'identifier plus de 2 000 régions différenciellement méthylées (DMR) subissant des changements de méthylation entre les états non-CSC et CSC radio-induit. Nous avons retenu 35 DMR présentant un profil de méthylation cohérent avec une potentielle contribution à la différenciation radio-induite. Cinq d'entre elles, associées aux gènes FSCN1, CHRNA6, CDH7, CD9 et PRKAR1B, ont été sélectionnées pour validation complémentaire. Les gènes régulés par ces changements de méthylation pourraient servir de nouvelles cibles thérapeutiques afin d'inhiber spécifiquement la conversion phénotypique de non-CSC à CSC et prévenir un enrichissement de la tumeur en CSC, réduisant ainsi le risque de rechute du cancer.

Pour valider les différences de méthylation observées en RRBS, la méthode de *Bisulfite Sequencing PCR* (BSP) a été choisie pour son accessibilité et son efficacité à quantifier les niveaux de méthylation d'un locus spécifique. En raison de l'absence d'outils à ce jour permettant d'analyser efficacement et de manière automatisée les résultats de BSP, provenant des deux approches de BSP (*direct-BSP* et *cloning-BSP*), nous avons donc fait le choix de développer sous R un nouvel outil, ABSP pour *Analysis of Bisulfite Sequencing PCR*. ABSP fournit une analyse complète, automatisée et accessible pour calculer les pourcentages de méthylation et comparer les différences de méthylation entre échantillons. ABSP et ses données associées sont téléchargeables à l'adresse <https://github.com/ABSP-methylation-tool/ABSP>. Ainsi, ce travail a mis en lumière l'importance de la méthylation de l'ADN dans la plasticité du phénotype souche cancéreux et le potentiel d'amélioration des outils d'analyse.

MOTS CLÉS : *cancer du sein, radiothérapie, cellules souches cancéreuses, méthylation de l'ADN, séquençage bisulfite, langage R*