



HAL
open science

Epigénétique et cancer : mécanismes et conséquences de l'activation anormale des gènes Cancer / Testis

Marthe Laisné

► To cite this version:

Marthe Laisné. Epigénétique et cancer : mécanismes et conséquences de l'activation anormale des gènes Cancer / Testis. Cancer. Université Paris Cité, 2021. Français. NNT : 2021UNIP7304 . tel-04084867

HAL Id: tel-04084867

<https://theses.hal.science/tel-04084867>

Submitted on 28 Apr 2023

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 Paris

Ecole doctorale Hématologie, Oncologie, Biothérapies (ED561)

Laboratoire «Epigénétique et Destin cellulaire»

ÉPIGÉNÉTIQUE ET CANCER :

MÉCANISMES ET CONSÉQUENCES DE L'ACTIVATION ANORMALE DES GÈNES CANCER/TESTIS

Par Marthe Laisné

Thèse de doctorat de Biologie

Dirigée par Pierre-Antoine Defossez

Présentée et soutenue publiquement le 11 octobre 2021

Devant un jury composé de :

Dr. Saadi Khochbin, IAB Grenoble	Rapporteur
Dr. Céline Vallot, Institut Curie Paris	Rapporteuse
Dr. Christophe Ginestier, CRCM Marseille	Examineur
Dr. Véronique Maguer-Satta, CRCL Lyon	Examinatrice
Dr. Raphaël Margueron, Institut Curie Paris	Examineur
Dr. Pierre-Antoine Defossez, UP Paris	Directeur de thèse

RESUME

Dans les cellules différenciées, les propriétés spécifiques à chaque type cellulaire sont assurées par la régulation spatio-temporelle de l'expression des gènes, qui repose notamment sur la combinaison de facteurs de transcription et de marques épigénétiques précisément distribuées. Une dérégulation de ces mécanismes peut entraîner la perte de l'identité cellulaire, et accompagne systématiquement le développement de certaines maladies comme le cancer. En effet, les cellules tumorales présentent de nombreuses anomalies épigénétiques et signalétiques qui modifient leur répertoire d'expression des gènes. En particulier, les tumeurs expriment fréquemment des gènes normalement restreints aux cellules germinales, appelés gènes Cancer/Testis. Certains gènes Cancer/Testis participent directement au processus tumorigénique. De plus, cette activation atypique peut être utilisée pour stigmatiser les cellules tumorales et servir de biomarqueurs ou de cibles thérapeutiques, notamment pour le développement d'immunothérapies anti-cancéreuses.

Ce travail présente deux approches permettant l'identification des mécanismes associés à la régulation de l'expression des gènes Cancer/Testis et à leur implication dans la tumorigenèse. Dans la première approche, des cribles génétiques ont permis d'identifier trois nouveaux acteurs importants dans la régulation du gène Cancer/Testis ADAM12, anormalement exprimé dans les cancers et impliqué dans la promotion de la migration cellulaire et des processus métastatiques. La seconde approche a étudié plus spécifiquement les tumeurs du sein : une approche bioinformatique a permis de découvrir de nouveaux marqueurs tumoraux parmi les gènes Cancer/Testis, dont certains présentent un potentiel pronostic et prédictif de la réponse au traitement par chimiothérapie. L'étude plus approfondie des tumeurs du sein basal-like a révélé l'implication de deux gènes Cancer/Testis, HORMAD1 et CT83, dans l'agressivité de certaines de ces tumeurs. In vitro, la co-expression de ces deux gènes par des cellules épithéliales mammaire a un effet synergique, suggérant une coopération fonctionnelle de ces deux gènes dans le développement tumoral. Ces données suggèrent des pistes prometteuses pour la compréhension, le diagnostic et le traitement de ces cancers très agressifs.

Mots-clés : Epigénétique, Biomarqueurs, Gènes Cancer/Testis, Cancer du sein basal-like

SUMMARY

In differentiated cells, the specific properties of each cell type are ensured by the spatio-temporal regulation of gene expression, which relies among other things on the combination of transcription factors and precisely distributed epigenetic marks. A deregulation of these mechanisms can lead to the loss of cellular identity, and systematically accompanies the development of certain diseases such as cancer. Indeed, tumour cells exhibit numerous epigenetic and signaling abnormalities that alter their gene expression repertoire. In particular, tumours frequently express genes normally restricted to germ cells, called Cancer/Testis genes. Some Cancer/Testis genes are directly involved in the tumorigenic process. Moreover, this atypical activation can be used to discriminate tumour cells and serve as biomarkers or therapeutic targets, especially for the development of anti-cancer immunotherapies.

This work presents two approaches to identify the mechanisms associated with the regulation of Cancer/Testis gene expression and their involvement in tumorigenesis. In the first approach, genetic screens were used to identify three important new players in the regulation of the Cancer/Testis gene ADAM12, which is abnormally expressed in cancers and involved in the promotion of cell migration and metastatic processes. The second approach studied breast tumours more extensively: a bioinformatics approach led to the discovery of new tumour markers among Cancer/Testis genes, some of which have prognostic and predictive potential for response to chemotherapy treatment. Further study of basal-like breast tumours revealed the involvement of two Cancer/Testis genes, *HORMAD1* and *CT83*, in the aggressiveness of some of these tumours. *In vitro*, co-expression of these two genes by mammary epithelial cells has a synergistic effect, suggesting a functional cooperation of these two genes in tumour development. These data suggest promising avenues for the understanding, diagnosis and treatment of these highly aggressive cancers.

Keywords: Epigenetics, Biomarkers, Cancer/Testis genes, Basal-like breast cancer

ACTIVATION DES GÈNES CANCER/TESTIS :
CAUSES ÉPIGÉNÉTIQUES ET CONSÉQUENCES
FONCTIONNELLES

Remerciements

Dans le discours qu'aujourd'hui je dois tenir, [...] j'aurais voulu pouvoir me glisser subrepticement. Plutôt que de prendre la parole, j'aurais voulu être enveloppé par elle, et porté bien au-delà de tout commencement possible. J'aurais aimé m'apercevoir qu'au moment de parler une voix sans nom me précédait depuis longtemps : il m'aurait suffi alors d'enchaîner, de poursuivre la phrase, de me loger, sans qu'on y prenne bien garde, dans ses interstices, comme si elle m'avait fait signe en se tenant, un instant, en suspens. [...] J'aurais aimé qu'il y ait derrière moi (ayant pris depuis bien longtemps la parole, doublant à l'avance tout ce que je vais dire) une voix qui parlerait ainsi: «Il faut continuer, je ne peux pas continuer, il faut continuer, il faut dire des mots tant qu'il y en a, il faut les dire jusqu'à ce qu'ils me trouvent, jusqu'à ce qu'ils me disent - étrange peine, étrange faute, il faut continuer, c'est peut-être déjà fait, ils m'ont peut-être déjà dit, ils m'ont peut-être porté jusqu'au seuil de mon histoire, devant la porte qui s'ouvre sur mon histoire, ça m'étonnerait si elle s'ouvre.»

M. Foucault, *L'ordre du discours* (1971)

Quelle inquiétude, et quelle excitation au moment d'écrire ces premières lignes. Et pourtant, en commençant, ou plutôt en finissant par le début ce manuscrit de thèse, comment ne pas essayer d'exprimer la gratitude et l'émotion que j'éprouve pour tous ceux et celles qui ont contribué à l'écrire.

Merci aux membres de mon jury d'avoir accepté d'écouter mon travail. De loin ou parfois de très près, vous avez inspiré cette thèse ; c'est un honneur que de l'achever avec vous. Merci en particulier pour votre superbe patience vis-à-vis des turpitudes administratives que nous avons rencontrées cet été.

Pierre-Antoine, je te dois toute cette thèse. Tu m'as aidé à affirmer mon ambition scientifique et mon désir de me lancer dans un doctorat de biologie, et j'ai appris au cours de ces années bien plus que je n'aurais osé l'espérer. Merci pour toutes nos discussions, tes conseils, ton écoute indéfectible, ces temps de réflexions scientifiques et humaines si précieux.

L'équipe, le lab, vous tous au quotidien : bordel de merci !! Laure, Nathaliya, Olivier, Kosuke, Lounis et Xiaoying à qui je souhaite le meilleur dans la poursuite de leurs doctorats, Ikrame, Mathieu, Nikhil, Cécilia, Sarah que j'espère comblés par leurs nouvelles aventures, tous les membres de cette si chouette unité : quel bonheur de croiser vos bonjours si souriants dans les couloirs (même masqués !), quelle chance que d'avoir vos conseils scientifiques quand « ça marche pâââs », et ces petits moments de vie en pause-café qui font tant de bien ! Merci pour votre gentillesse, votre chaleur et votre soutien, ça été fondamental.

Les batailles : Samantha, Jennifer, Maxime, Sandra, Christelle, Agnèle, Droopy, le Gros Poisson, sans qui tout ça m'aurait tellement coûté, et grâce à qui tout ça a été tellement plus doux. Merci.

Evidemment, les filles, Maddy Jo Ju Clara No, un magnifique collier de syllabes essentielles. Dieu comme je vous aime ! J'ai tant de chance de vous avoir dans ma vie. Merci pour ces cafés-debriefs du jeudi, du lundi, et du mercredi aussi, éclat de tendresse dans les semaines-tunnels, pour les Brunchs & Yoga qui se transformaient en chips, houmous & papotte, pour la confiance et la force que je puise en vous.

Ma coloc étendue chérie : Elena, Margot, Thomas, Bernard-Florent, Benjamin, merci pour toutes nos escapades, nos bêtises, nos tendres vidages de sac et vidange de «JPP», pour les plans sur la comète et les week-ends qui passent en une éclipse. Ca y est, on est tous toutes Docteur-es !!!! Fièrè :)

Emile, Agathe, mes parents, Jacqueline, Ghislaine, et puis encore Benoit et Léone, tous chacun et particulièrement j'ai écrit pour vous, à un moment donné. Vous êtes présents dans mon cœur chaque fois que j'entreprends quelque chose qui compte – et aussi, quand je fais des pâtes au pesto -.

Edith, mon amour. Merci c'est trop petit, alors je nous souhaite de continuer à grandir, à explorer, à construire nos vies ensemble, à vieillir et à aimer toujours, et à faire plein de petits et grands articles.

Tous, vous êtes mes racines et mes ailes (oui, sérieusement). Merci !!

INTRODUCTION

CHAPITRE 1 - Epigénétique et destin cellulaire**9****Partie I - À quoi ressemble le paysage épigénétique des mammifères ?**

- | | | |
|----|---|-----|
| 1. | Structure d'un gène codant pour une protéine | 10 |
| 2. | Organisation de la chromatine | 12 |
| 3. | La méthylation de l'ADN : modification épigénétique des acides nucléiques | 13 |
| 4. | Les modifications des histones : Modifications épigénétiques de la chromatine | 14 |
| 5. | Couplage des états de méthylation et des marques d'histones | 120 |

Partie II - Comment les marques épigénétiques sont-elles mises en place, propagées, interprétées et éventuellement remodelées ?

- | | | |
|----|---|----|
| 1. | Lire, écrire et éditer les épigénomes : fonctions des Readers, Writers et Erasers | 23 |
| 2. | Coordination épigénétique grâce à l'association des acteurs chromatiniens | 25 |
| 3. | Mise en place de novo des marques épigénétiques lors du développement | 27 |
| 4. | Maintien des marques épigénétiques à travers les divisions cellulaires | 29 |
| 5. | Propagation épigénétique des signaux dans le temps | 31 |

Partie 3 - Quelles sont les fonctions essentielles des marques épigénétiques ?

- | | | |
|----|---|----|
| 1. | Restreindre le répertoire d'expression des gènes | 34 |
| 2. | Protéger l'information génétique et orchestrer sa duplication | 38 |
| 3. | Altération des marques épigénétiques et pathologies humaines | 40 |

CHAPITRE 2 - Cancers et soi-modifié**43****Partie I - Quelles sont les caractéristiques fondamentales et permissives des cancers ?**

- | | | |
|----|---|----|
| 1. | Huit caractéristiques fondamentales caractérisent tous les cancers | 45 |
| 2. | Deux caractéristiques permissives favorisent le développement oncogénique | 49 |
| 3. | Les tumeurs sont des systèmes hétérogènes | 50 |
| 4. | Les consortia de génétique et épigénétique des cancers | 55 |

Partie II - Quel est le rôle de l'épigénétique dans la progression oncogénique ?

- | | | |
|----|---|----|
| 1. | Les altérations épigénétiques participent à la transformation | 58 |
| 2. | L'altération locale des profils épigénétiques modifie l'expression des gènes | 61 |
| 3. | La plasticité des états tumoraux est soutenue par des transitions épigénétiques | 63 |

Partie III - Les gènes Cancer/Testis sont-ils un modèle crédible pour investiguer les altérations épigénétiques des cellules cancéreuses ?

- | | | |
|----|--|----|
| 1. | Des gènes épigénétiquement réservés à la lignée germinale mâle | 65 |
| 2. | Des gènes anormalement exprimés par de nombreuses tumeurs | 67 |
| 3. | Les gènes C/T sont un objet d'étude prometteur en oncologie | 69 |

Partie I - Quelle est l'origine des cancers du sein ?

1. Epidémiologie et facteurs de risque de cancers du sein 72
2. La glande mammaire présente une structure hiérarchique 73
3. Processus de transformation carcinomateuse 75

Partie II - Comment classer les cancers du sein ?

1. Classifications histologiques 77
2. Classification sur la base de l'expression des récepteurs hormonaux et de HER2 77
3. Classifications moléculaires 79

Partie III - Quelles sont les propriétés des sous-types de cancers du sein ?

1. Les cancers du sein ont différentes origines cellulaires et moléculaires 84
2. L'instabilité génomique est modérée (Luminal A), à forte (Basal-Like) 85
3. Les réseaux de régulation transcriptionnels miment ceux de la cellule d'origine 86
4. Des facteurs épigénétiques impliqués dans la différenciation des cancers du sein 87
5. Une hétérogénéité intratumorale variable en fonction du sous-type 89
6. Les gènes Cancer/Testis dans les cancers du sein 92

Objectifs de la thèse**93****RESULTATS****95****Partie I. Régulation de l'expression des gènes C/T dans les cellules non transformées**

1. Définition du cadre expérimental 96
2. Conclusions majeures du projet 97
3. Résultats supplémentaires: vers l'étude des cancers 129

Partie II. Régulation de l'expression des gènes C/T dans les cancers du sein

1. Définition du cadre expérimental 134
2. Résultats supplémentaires 187

DISCUSSION**195**

- Résumé des résultats 196
- Quelles sont les mécanismes d'activation des gènes C/T ? 198
- Comment expliquer les fonctions émergentes de la co-activation d'HORMAD1 et de CT83 ? 200
- Quel est la chronologie des événements d'activation des gènes C/T et de transformation ? 203
- Vers des immunothérapies visant les antigènes C/T ? 205

ANNEXES - Autres contributions scientifiques**207****BIBLIOGRAPHIE****229**

Table des figures

INTRODUCTION

Figure 1 : Composition du génome humain	10
Figure 2 : Lecture et expression d'un gène codant pour une protéine	11
Figure 3 : Chromatine(s) et expression des gènes	12
Figure 4 : Organisation de la chromatine : de la séquence nucléotidique aux nucléosomes	13
Figure 5 : Organisation de la chromatine : compaction et repliement 3D	14
Figure 6 : Conséquences moléculaires de la méthylation des cytosines chez les mammifères	15
Figure 7 : Distribution génomique des cytosines méthylables	17
Figure 8 : Modifications post-traductionnelles des histones	19
Figure 9 : Combinaisons des différents niveaux de régulation épigénétique	21
Figure 10 : Writers, Readers & Erasers	23
Figure 11 : Writers, Readers & Erasers 2	24
Figure 12 : Coopération des acteurs épigénétiques lors de répression transcriptionnelle	25
Figure 13 : Coopération des fonctions chromatinienne	26
Figure 14 : Développement embryonnaire et différenciation cellulaire	27
Figure 15 : Divergence précoce des lignées germinales et somatiques	28
Figure 16 : Cinétique de la restauration des modifications d'histones au cours du cycle cellulaire	29
Figure 17 : Division cellulaire et réplication de la chromatine	30
Figure 18 : Potentialisation épigénétique des signaux cellulaires	32
Figure 19 : Classification de tissus humains à partir de données de méthylation d'ADN	35
Figure 20 : Classification des gènes codants selon leur spécificité d'expression	37
Figure 21 : Cinétique de la réplication et structure de l'information génétique	42
Figure 22 : Caractéristiques fondamentales et permissives des cancers	44
Figure 23 : Inactivation des gènes suppresseurs de tumeurs : TP53, un master régulateur	46
Figure 24 : Etats de transition de l'EMT et progression métastatique des cellules carcinomateuses	48
Figure 25 : Evolution conceptuelle de la notion de tumeur	51
Figure 26 : Evolution clonale des tumeurs	53
Figure 27 : Changements de la structure chromatinienne des cellules cancéreuses	58
Figure 28 : Cinétique des anomalies épigénétiques au cours de la progression tumorale	59
Figure 29 : Anomalies de la méthylation de l'ADN et dérégulation de l'expression des gènes	62
Figure 30 : Expression des gènes C/T au cours de la différenciation des gamètes mâles	66
Figure 31 : Proportions de gènes tissus-restreints anormalement activés par les tumeurs, en fonction de leur origine	67
Figure 32 : Expression des gènes C/T	68
Figure 33 : Les gènes C/T peuvent agir en tant qu'oncogènes	69
Figure 34 : Evolution des cancers du sein entre 1980 et 2012	72
Figure 35 : Classification anatomopathologique des cancers canaux	75

Figure 36 : Sous-types de cellules épithéliales de la glande mammaire normale	75
Figure 37 : Exemples de marquage immunohistochimiques (IHC)	77
Figure 38 : Identification des sous-types moléculaires de cancers du sein	79
Figure 39 : Correspondance entre les différentes classifications	80
Figure 40 : Evolution des classifications moléculaires des tumeurs triple-négatives	82
Figure 41 : Deux modèles pour expliquer la diversité des sous-types de cancers	84
Figure 42 : Réseaux de régulation des sous-types cellulaires mammaires	86
Figure 43 : Facteurs épigénétiques impliqués dans la différenciation de la glande mammaire	88
Figure 44 : Variation des proportions de cellules souches cancéreuses et de leurs descendances différenciées selon le sous-type de tumeurs du sein	90
Figure 45 : Hétérogénéité génétique et épigénétique dans les cancers du sein	91

RESULTATS - I

Figure 1 : Réponses transcriptomiques des IMR90 à deux drogues épigénétiques	129
Figure 2 : Représentation schématique des valeurs seuils d'activation des gènes C/T	129
Figure 3 : Activation des 42 gènes C/T sélectionnés pour cette étude dans 4 types de cancers	120
Figure 4 : Paysage épigénétique de 4 gènes dans les IMR90	132

RESULTATS - II

Figure 1 : Modèle cellulaire de tumorigenèse de cellules épithéliales mammaires	135
Figure 2 : Phénotype des sous-populations de HMLE par observation microscopique	136
Figure 3 : Phénotype des sous-populations de HMLE par cytométrie	137
Figure 4 : Premières analyses fonctionnelles du rôle de HORMAD1 et CT83 dans la lignée HMLE	188
Figure 5 : Premières analyses fonctionnelles du rôle de HORMAD1 et CT83 dans la lignée HME	189
Figure 6 : Piste pour la découverte des mécanismes d'activation d'HORMAD1 et de CT83	191
Figure 7 : Influence de la cellule d'origine pour l'activation d'HORMAD1 et de CT83	193

DISCUSSION

Figure 1 : Activation des gènes C/T et transformation tumorale	196
Figure 2 : Bilan des résultats obtenus par nos deux approches expérimentales	197
Figure 3 : Facteurs pouvant influencer la probabilité d'activation d'HORMAD1 et de CT83	199
Figure 4 : Modélisation des hypothèses à l'origine de l'émergence de nouvelles propriétés suite à la co-activation d'HORMAD1 et de CT83	202
Figure 5 : Deux modèles chronologiques pour l'activation d'HORMAD1 et de CT83	204

CHAPITRE 1

Epigénétique et destin cellulaire

Partie I

À quoi ressemble le paysage épigénétique des mammifères ?

La section 1 de ce chapitre présentera un aperçu global des structures épigénétiques. La **section 2** s'intéressera aux mécanismes permettant la mise en place et le maintien de ces états épigénétiques ; enfin la **section 3** discutera des différents rôles biologiques des modifications épigénétiques. Dans la suite de ce texte, seul le cas des mammifères sera abordé et plus précisément celui du génome humain : en effet les mécanismes épigénétiques diffèrent largement entre clades, et ne sont donc pas transposables d'un organisme à l'autre ; d'autre part les modèles humains et murins ont été largement documentés dans le cadre des études épigénétiques, et sont ceux auxquels mon travail de thèse s'est intéressé.

1. Structure d'un gène codant pour une protéine :

A l'échelle moléculaire, un gène est défini comme **une information contenue dans une séquence de nucléotides** (tel que démontré par Avery en 1944), **susceptible de s'exprimer et déterminant un caractère héréditaire**. On compte entre 20 000 et 35 000 gènes dans le génome humain (selon les définitions), répertoriés dans des bases de données telles que ENSEMBL (Cunningham et al. 2015). Les gènes peuvent être transcrits en ARN (environ 60% du génome humain est effectivement transcrit) : on distinguera alors les **ARN codants** pour une protéine fonctionnelle, et les **ARNs non-codants** (FIGURE 1). Ces derniers ne sont pas traduits et peuvent être des ARNs structuraux (ARN ribosomiaux et de transferts) ou des ARNs fonctionnels importants pour la régulation de l'expression du génome (lncRNA, miRNA, piRNA, snoRNA, snRNA...). Les segments non codants du génome (FIGURE 1) jouent un rôle essentiel dans la régulation fine de la synthèse des protéines, mais également en tant qu'éléments architecturaux : en tant que tels, ils sont aussi responsables d'une information héréditaire, complexifiant la définition de gène.

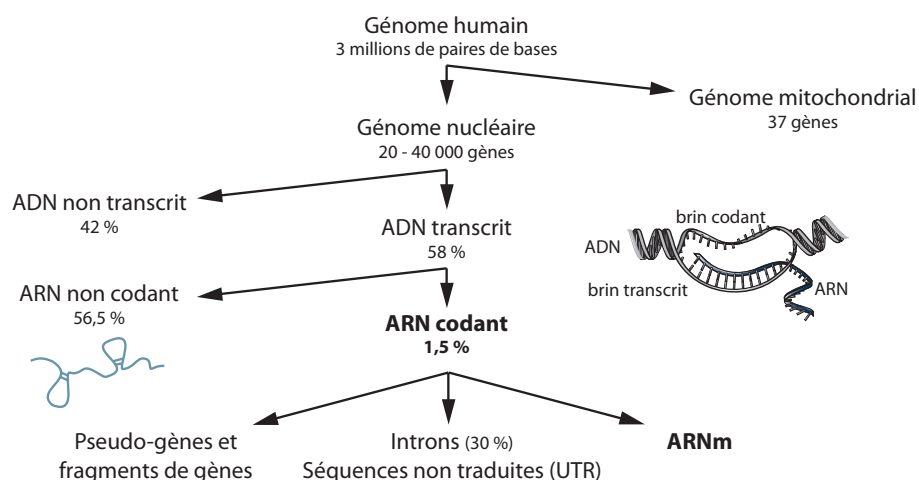


Figure 1 : Composition du génome humain

Les proportions des différents types de séquences sont indiquées.

Dans le cadre de mon travail de thèse je me suis principalement intéressée aux gènes codants, dont la structure est présentée dans la **FIGURE 2**. On distingue deux types de séquences composant le gène : le cadre ouvert de lecture ; et la région de contrôle.

Le **cadre ouvert de lecture** est défini en 5' par le site d'initiation de la transcription (TSS : *Transcription Start Site*) et en 3' par des séquences signal indiquant la fin de transcription, tels que le signal de polyadénylation AATAAA. Au sein de cette région transcrite, on distingue des séquences appelées introns et exons. Les séquences introniques transcrites seront éliminées par épissage lors de la maturation de l'ARNm. On trouve également dans l'ARNm mature, en amont du codon *start* AUG et en aval du codon *stop*, des séquences non traduites par les ribosomes appelées séquence UTR (de l'anglais *Untranslated*), importantes pour la stabilité de l'ARN.

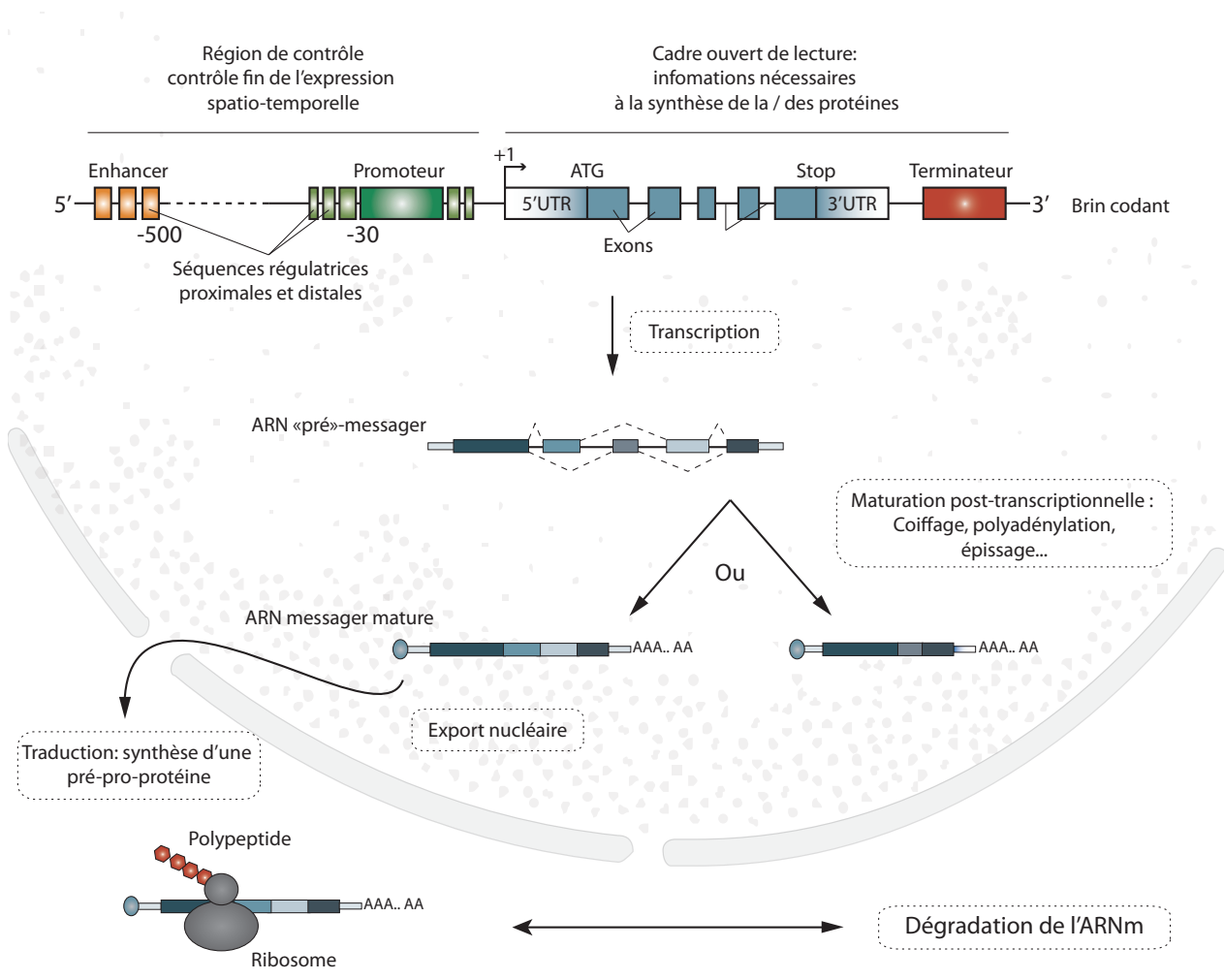


Figure 2 : Lecture et expression d'un gène codant pour une protéine

Tous ces mécanismes ne sont pas épigénétiques, mais contribuent à l'activité des gènes. A ceux représentés ici, il faudrait également ajouter les étapes contrôlant la maturation, l'adressage et la dégradation des protéines.

La **région de contrôle** comporte une **région centrale** (ou promoteur minimal) d'environ 100 paires de bases (pb), à laquelle se fixe l'ARN polymérase et les facteurs généraux de la transcription. La région centrale est souvent flanquée de **séquences promotrices** liant des facteurs de transcription spécifiques. Enfin, certains gènes possèdent parfois des séquences éloignées de la région promotrice, à plusieurs kilobases de distance du gène voire sur un autre chromosome, et qui augmentent (**enhancer**) ou diminuent (**silencers**) l'activité du complexe transcriptionnel grâce au repliement tridimensionnel de l'ADN dans le noyau.

La combinaison des séquences promotrices et des enhancers/silencers est à l'origine de la diversité de régions de contrôle, essentielle à la régulation fine de l'expression spatio-temporelle des gènes. Pour assurer le renforcement de ces patrons d'expression au cours des divisions cellulaires, ces éléments sont l'objet de modifications épigénétiques qui permettent la propagation de l'information d'une génération cellulaire à la suivante. Ces modifications épigénétiques ne sont pas encore toutes élucidées, mais leur découverte a initié une vraie révolution dans la compréhension des mécanismes de régulation de l'expression des gènes, et a propulsé au-devant de la scène le système chromatinien, que nous allons à présent décrire.

2. Organisation de la chromatine :

Au sein du noyau, la molécule d'ADN n'est pas nue : bien au contraire l'ADN est associé à de nombreuses protéines et s'organise en une structure nucléoprotéique appelée **chromatine**. La compaction de la chromatine constitue une barrière physique pour les événements faisant intervenir l'information génétique (comme la réplication, la réparation ou la transcription), qui nécessitent d'avoir accès à la séquence d'ADN. C'est pourquoi le véritable théâtre de ces événements est bien la chromatine et son organisation tridimensionnelle, organisant et régulant l'accessibilité de l'ADN.

Historiquement, la chromatine fut divisée en deux domaines distincts à partir d'observations au microscope électronique (**FIGURE 3A**):

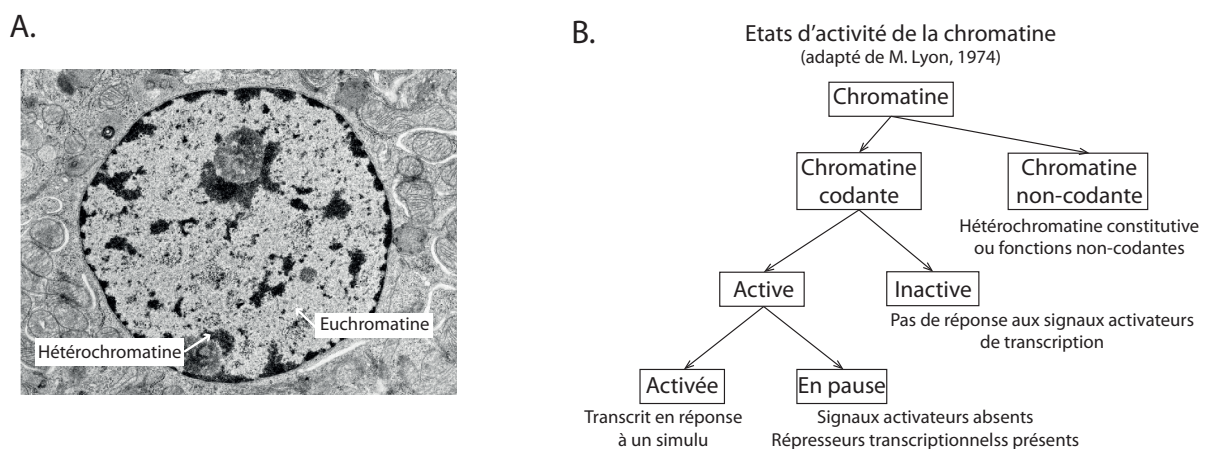


Figure 3 : Chromatine(s) et expression des gènes

- A. Photographie d'un noyau cellulaire vu en microscopie électronique.
B. Différents états d'activité de la chromatine (Adapté de Lyon 1974)

l'hétérochromatine localisée en périphérie du noyau et très dense aux électrons, correspondant à des régions compactes de la chromatine, qui s'oppose à **l'euchromatine**, plus claire et décondensée au cours de l'interphase (Heitz et al. 1928). L'hétérochromatine est elle-même divisée en deux sous-catégories : **l'hétérochromatine constitutive**, qui reste condensé dans tous les types cellulaires et qui contient peu de gènes, formées principalement des séquences répétées situées à proximité des centromères et des télomères ; et **l'hétérochromatine facultative** qui contient des régions codantes mises sous silence dans certains types cellulaires ou à certains stades du développement. **La définition précise des régions condensées d'hétérochromatine facultative est cruciale pour assurer l'identité des cellules et sa transmission à travers les divisions cellulaires.** Ces régions plus ou moins condensées de la chromatine sont le résultat d'une organisation hiérarchique, dont les différentes couches organisationnelles se surimposent les unes aux autres (Bernstein et al. 2011).

A la racine, on trouve la **séquence primaire de l'ADN**, dont les nucléotides peuvent être l'objet de **modifications chimiques directes** (FIGURE 4A) : en particulier la méthylation des cytosines en 5' en contexte CpG, qui joue un rôle dans l'expression des gènes (Sinsheimer et al. 1955). La double hélice d'ADN s'enroule ensuite autour d'un complexe de 8 molécules d'histones centrales (deux copies des histones H2A, H2B, H3 et H4), le tout stabilisé par l'histone H1 (FIGURE 4B). Ce complexe protéine-ADN forme le **nucléosome**, unité de base de la chromatine (Kornerg & Thomas, 1974). Les queues N-terminales des histones centrales sont flexibles, émergent du domaine globulaire, et font l'objet de nombreuses **modifications post-traductionnelles** (FIGURE 4C) pouvant affecter la condensation et le niveau de transcription de la chromatine (Zhou, Bannister & Kouzarides, 2011). Enfin, les histones de cœurs peuvent être échangées par des variants d'histones, qui auront un effet sur la stabilité et les interactions des nucléosomes.

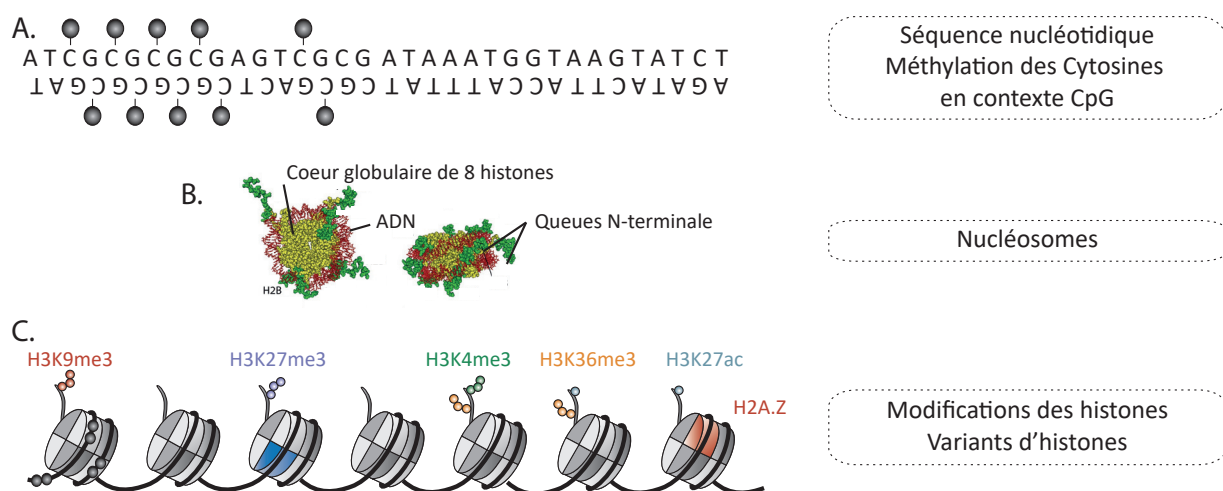


Figure 4 : Organisation de la chromatine : de la séquence nucléotidique aux nucléosomes

- A. Modification chimique de la séquence nucléotidique d'ADN. Les groupes méthyles sont représentés par des épingles noires.
- B. Nucléosome en représentation 3D. Cette structure est composée d'un octamère d'histones autour duquel s'enroule la molécule d'ADN, stabilisé par l'histone H1. Les queues N-terminales émergent du domaine globulaire.
- C. Répétitions de nucléosomes et exemples choisis de modifications chimiques et variants d'histones.

La structure primaire de la chromatine est déterminée par la position des nucléosomes, ainsi que les modifications et variants d'histones. Celle-ci peut être organisée de façon assez souple et décompactée, sous forme de « **collier de perles** » tel que décrit par Olins et Olins en 1974 (**FIGURE 5A**) ; s'enrouler sur elle-même grâce à l'empaquetage des nucléosomes en une structure cylindrique compacte de 30nm ; ou se compacter davantage en se repliant en **boucles** le long d'une matrice protéique centrale (la matrice nucléaire, **FIGURE 5B**). Ces boucles de chromatine ne sont pas dispersées dans l'ensemble du noyau : au contraire, chaque chromosome occupe un territoire nucléaire bien précis (on parle de **territoire chromosomique**). Cette régionalisation joue un rôle important dans la régulation de l'expression des gènes : les différents territoires sont agencés de telle sorte que les régions riches en gènes soient plutôt au centre du noyau, tandis que les régions plus pauvres en gènes sont plutôt déportées en périphérie (**FIGURE 5B**). C'est pourquoi la conformation tridimensionnelle de la chromatine définit des domaines de régulation de l'expression des gènes, pouvant être activement transcrits (tels que les **centres de transcription**) ou majoritairement réprimés (comme les **domaines associés à la lamina**).

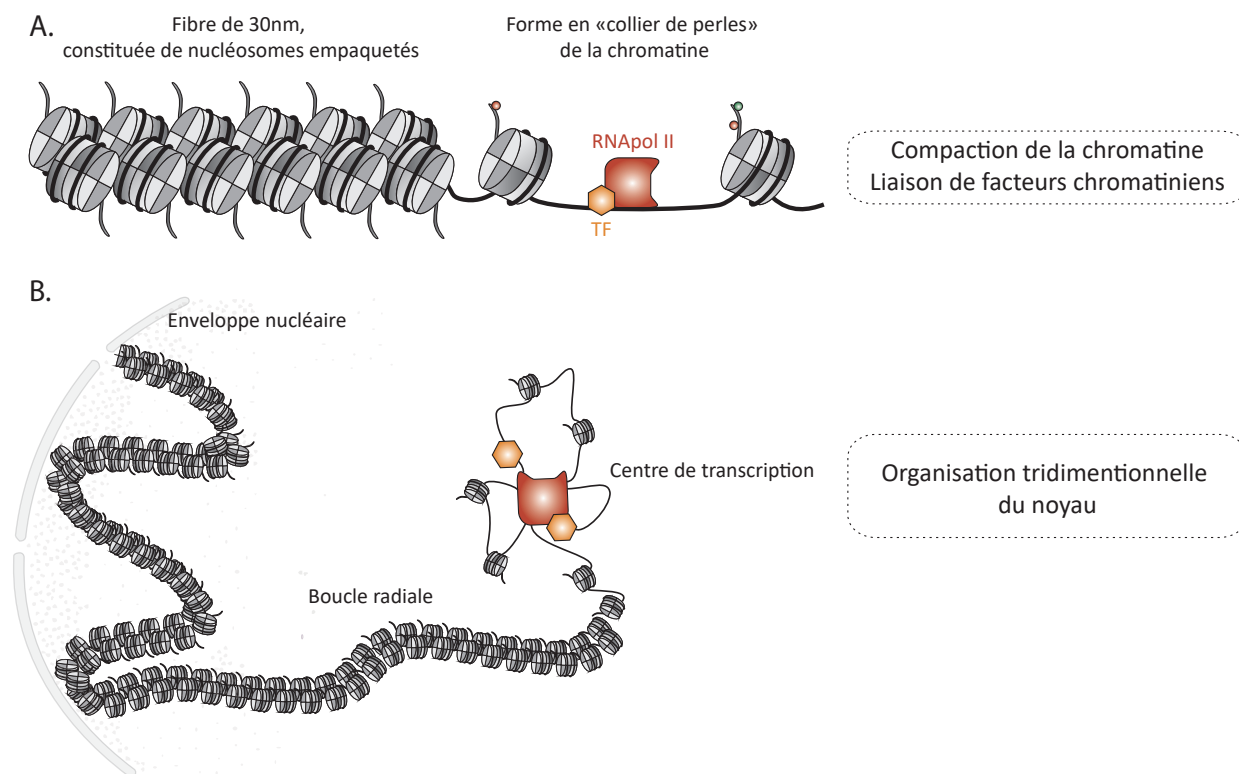


Figure 5 : Organisation de la chromatine : compaction et repliement 3D

A. Etats de compaction de la chromatine et accessibilité de la molécule d'ADN

B. Organisation tridimensionnelle de la chromatine et régionalisation fonctionnelle du noyau (Adapté de Zhou 2011)

Etant donné que la structure hiérarchique de la chromatine repose sur les niveaux d'organisation inférieurs, nous allons à présent décrire plus précisément ces marques épigénétiques cruciales que sont les modifications d'histones et la méthylation de l'ADN.

3. La méthylation de l'ADN : modification épigénétique des acides nucléiques

L'ADN des mammifères peut être modifié de façon covalente par méthylation du carbone 5 des cytosines (abrégié en 5mC) ; les 5mC se trouvent quasi-exclusivement dans le contexte de dinucléotides 5'-CpG-3' (abrégié en CpG). La méthylation de l'ADN est une **modification absolument essentielle** au développement des mammifères : les mutants murins déficients pour cette modification présentent des anomalies du développement très sévères, conduisant à une mortalité embryonnaire précoce (Li et al. 1992, Okano et al. 1999)

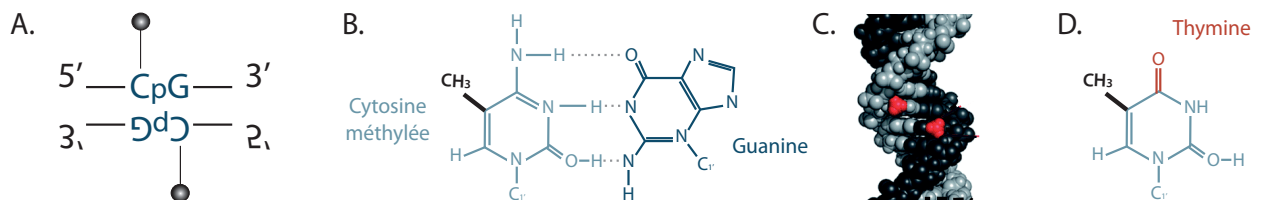


Figure 6 : Conséquences moléculaires de la méthylation des cytosines chez les mammifères

A.-C Représentation schématique, moléculaire et 3D de la méthylation des cytosines en contexte palindromique CpG.

D. Produit de la déamination spontanée de la 5-méthyl-cytosine.

La marque 5mC est généralement présente symétriquement sur les deux brins (**FIGURE 6A**). La configuration moléculaire de l'ADN est telle que l'ajout de ce groupe méthyl sur une cytosine ne perturbe pas les liaisons hydrogènes avec sa base complémentaire, la guanine (**FIGURE 6B**). De plus, l'encombrement stérique occasionné par l'ajout d'un groupement méthyl sur les CpG est minimale : la structure 3D de la double hélice d'ADN n'est pas affectée par cette modification épigénétique (**FIGURE 6C**), ce qui pose la question de sa reconnaissance et de son interprétation par la cellule.

Dans les cellules somatiques humaines, **la base 5mC représente 1% de toutes les bases de l'ADN**, soit 70 à 80% des CpG qui se trouvent méthylés. Cependant, la répartition des CpG méthylés varie au cours du développement et selon les types cellulaires : la distribution des 5mC n'est pas, à première vue, aléatoire. Premier biais dans la répartition de la méthylation de l'ADN chez les mammifères : bien que d'autres nucléotides peuvent en théorie être méthylés, **la méthylation concerne principalement les cytosines en contexte CpG**.

Deuxième biais : **le dinucléotide CpG est très faiblement représenté** par rapport aux autres dinucléotides. Dans le génome humain, le contenu en GC est de 41%, donc la fréquence attendue de CpG devrait être de $0,21 \times 0,21 = 4\%$. Or, la fréquence observée de CpG est seulement de 0,8%, soit un cinquième de la fréquence attendue. En réalité, la méthylation des cytosines a un coût : le produit de la déamination de la 5-méthyl-cytosine est la thymine (**FIGURE 6D**), or cette base est légitime dans l'ADN. Il est donc plus complexe d'identifier et de réparer la 5mC déaminé ; souvent, cette altération sera confondue avec un mésappariement et conduira à une transition C->T, d'où le caractère intrinsèquement mutagène des 5mC et la faible proportion de CpG dans le génome, effacés au fil de l'évolution.

Le troisième biais concerne la répartition des CpG dans le génome : **ils sont souvent densément regroupés sur des régions de plusieurs centaines de bases appelées îlots CpG**, par opposition au reste du génome où les CpG sont plutôt rares (**FIGURE 7A**). La définition formelle d'un îlot CpG est plus ou moins restrictive selon les auteurs : un îlot CpG typique mesure généralement 1 à 2kb de longueur, et présente une fréquence de GC supérieure à 70% (contre 41% dans le reste du génome) avec une forte densité en dinucléotide CpG (CpG > 60%). Le génome humain compte environ 29 000 îlots CpG, préférentiellement associés avec les promoteurs et premiers exons des gènes. Ainsi, environ 60% des gènes du génome humain sont associés à un îlot CpG (**FIGURE 7B**). **Ces promoteurs très riches en CpG (High CpG density, HCP), sont souvent associés aux gènes de ménage (Brutlag et al. 2006, Weber et al. 2007)**. Cette distinction entre promoteurs riches ou pauvres en CpG aura des conséquences sur les mécanismes régulant leur activité. Enfin, la densité en CpG varie au sein même du cadre de lecture de la majorité des gènes (**FIGURE 7C**) : la quantité de CpG augmente drastiquement au niveau du promoteur, pour atteindre sa valeur maximale dans le premier exon. Elle est légèrement plus élevée dans les exons que dans les introns, et diminuera sensiblement au fil du dernier exon et de la région de terminaison de la transcription.

La majorité des promoteurs de types HCP restent non-méthylés quel que soit le stade de développement de l'organisme : **la méthylation du promoteur est associée à sa répression transcriptionnelle (Bird & Wolffe 1999)**, or les gènes de ménages, souvent associé à un îlot CpG, sont exprimés dans tous les types cellulaires. Au contraire, les CpG méthylés sont plutôt associés aux éléments transposables (qui doivent rester réprimés), aux régions intergéniques et au corps des gènes transcrits (**FIGURE 7D**) ; l'absence de méthylation de l'ADN corrèle donc avec les régions où la transcription peut s'initier.

La première étude de méthylation des CpG à l'échelle du génome humain a mis en évidence des **régions différenciellement méthylées selon le type cellulaire**, ce niveau de méthylation corrélant avec l'activité des gènes (Lister et al. 2009). A partir de ces résultats, plusieurs études ont cherché à segmenter le génome des différents tissus humains en fonction de leur méthylation. Les résultats obtenus montrent que la méthylation de l'ADN est l'un des marqueurs les plus discriminants pour identifier le tissu d'origine d'un échantillon : en effet, cette marque dessine une signature persistante, dont on retrouve la trace à travers les générations cellulaires. Cette trace robuste résiste en partie à des événements extrêmement déstabilisant, comme la transformation tumorale : on peut ainsi identifier de façon fiable le nid initial d'une métastase d'origine inconnue grâce à son patron de méthylation, qui demeure plus proche de son tissu d'origine que de ceux des autres tissus (Hoadley et al. 2018).

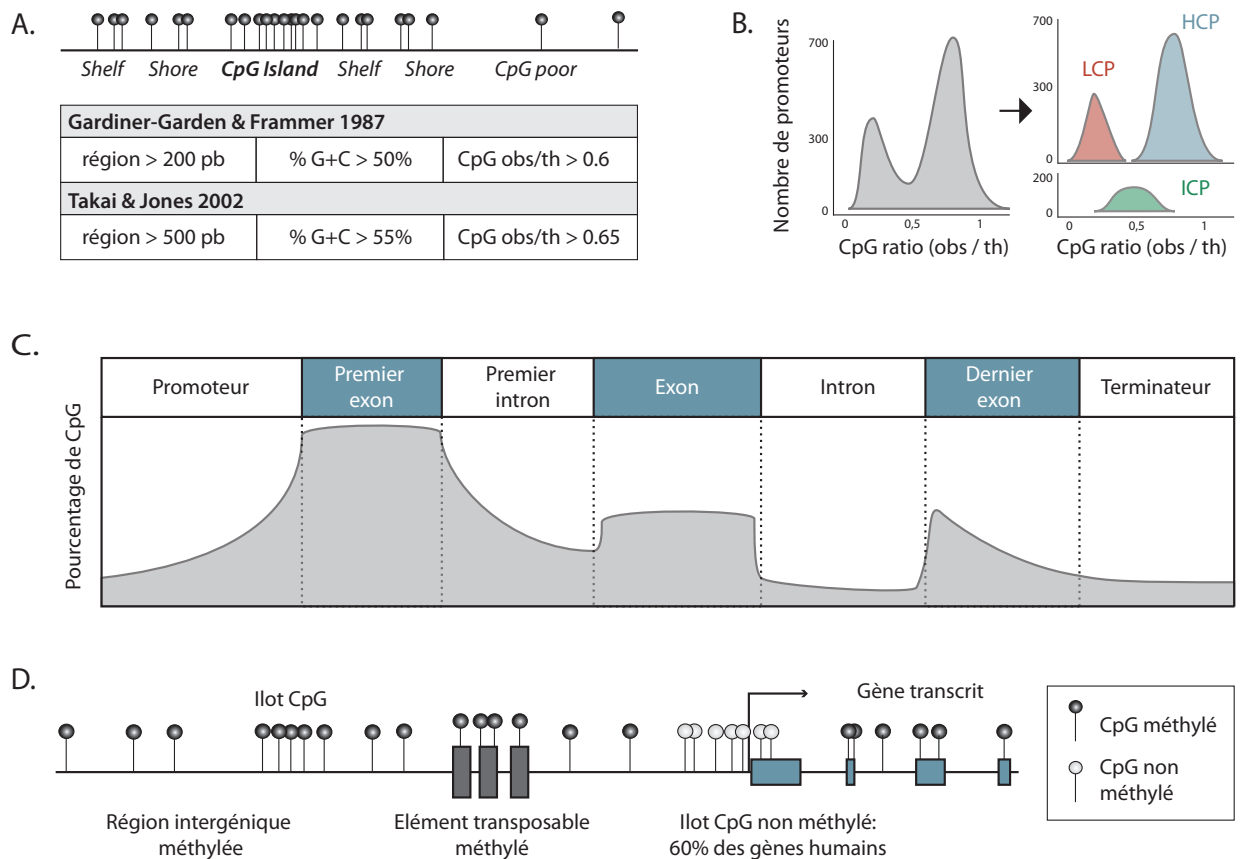


Figure 7 : Distribution génomique des cytosines méthylables

A. Nomenclature et définitions des îlots CpG et leurs régions flanquantes.

B Classification des promoteurs humains en fonction de leur densité en CpG (Adapté de Weber et al. 2007)

C. Représentation schématique d'un gène humain type et de sa densité en CpG.

D. Elements génomiques fonctionnels et densité en 5mC. Les CpG méthylés sont représentés par une épingle noire, non-méthylés par une épingle blanche.

4. Les modifications des histones : Modifications épigénétiques de la chromatine

Le deuxième niveau d'organisation de la chromatine implique les modifications post-traductionnelles des protéines histones. Ces modifications peuvent affecter une soixantaine d'acides aminés des quatre histones canoniques, dont les résidus lysine (K), arginine (R) et sérine (S) (**FIGURE 8A**). Un même acide aminé peut faire l'objet d'un très grand nombre de modifications distinctes : c'est par exemple le cas de la lysine 12 de l'histone 4, détectée sous 6 états différents (non modifiée, acétylée, biotinylée, mono-, di- ou tri-méthylée). Certaines modifications, telles que l'acétylation ou la phosphorylation, modifient la charge globale du nucléosome et donc les interactions électrostatiques qui déterminent la compaction de la chromatine (**Bowman & Poirier 2015**). De plus, les modifications des queues N-terminales des histones sont facilement accessibles et constituent des modules de reconnaissance pour les protéines nucléaires. Enfin, les variants d'histones peuvent avoir à la fois des effets directs sur les nucléosomes en entraînant la modification de l'environnement local de la chromatine, et des effets indirects en autorisant le recrutement spécifiques de partenaires (**Buschbeck & Hake, 2017 ; Talbert & Henikoff, 2016**). Certaines marques se font compétition et sont mutuellement exclusives, comme l'acétylation et la tri-méthylation de H3K27, de plus différentes modifications peuvent coexister sur une même protéine et sur un même nucléosome, créant une **combinatoire** aux multiples possibilités (**Bannister & Kouzarides 2011, Su & Denu 2016, FIGURE 8B**). Toutes ces observations ont conduit à formuler l'hypothèse d'un «code histone», à l'image du code génétique mais impactant la structure chromatinienne et son accessibilité (**Strahl & Allis, 2000**)

Le « code histone » corrèle avec l'état de compaction de la chromatine : certaines modifications post-traductionnelles d'histones ont été décrites comme caractéristiques de **l'hétérochromatine (H3K9me3, H3K27me3)**, d'autres de **l'euchromatine (H3K4me3, H3K9ac)**. On retrouve, comme pour la méthylation de l'ADN, une corrélation forte entre les domaines fonctionnels de l'ADN et les modifications des histones (**FIGURE 8C**) : les promoteurs actifs des gènes transcrits sont généralement décorés des marques H3K4me3 et H3K27ac, avec un enrichissement caractéristique de part et d'autre du TSS ; le corps des gènes transcrits présente deux gradients antagonistes de H3K79me3 et H3K26me3 ; les enhancers quant à eux combinent généralement un double pic d'H3K27ac et un enrichissement plus diffus d'H3K4me1. Le décryptage progressif de ce « code histone » a ainsi aidé à l'annotation fonctionnelle de la chromatine dans les différents types cellulaires, car ces marques jouent un rôle essentiel dans la régulation transcriptionnelle, la réparation, la réplication de l'ADN, l'épissage alternatif et la condensation (**Portela & Esteller 2010**).

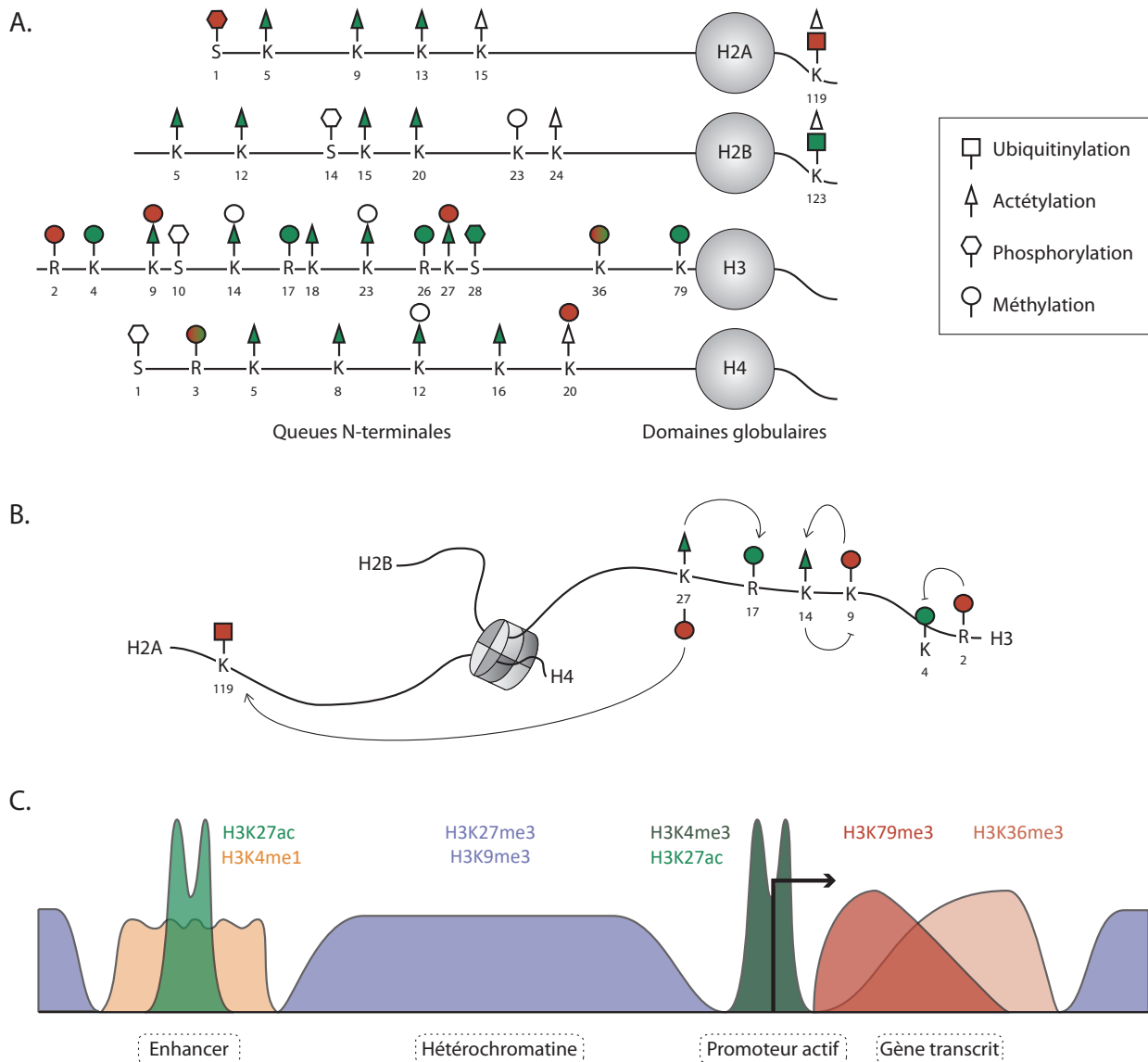


Figure 8 : Modifications post-traductionnelles des histones

A. Principales modifications post-traductionnelles des histones. Les modifications associées à un effet positif sur la transcription sont représentées en vert, celles ayant un effet négatif sont en rouge. Les marques vides indiquent que le lien entre ces modifications et la transcription n'a pas été étudié extensivement. (Adapté de Weinberg 2006, Rothbart & Strahl 2014)

B. Relations de renforcement ou d'exclusion de différentes modifications d'histones. Un effet positif est indiqué par une flèche, un effet négatif par une barre perpendiculaire (Adapté de Bannister & Kouzarides 2011)

C. Régionalisation fonctionnelle des modifications des histones.

5. Couplage des états de méthylation et des marques d'histones

D'importants projets de cartographie de la chromatine de tissus et cellules humaines ont été entrepris par le *NIH Roadmap Epigenomic Mapping Consortium* (Kundaje et al. 2015) et le projet *Encyclopedia of DNA Elements* (Davis et al. 2018), notamment. Associant différentes techniques de séquençages, ces projets ont produits des cartes précises des modifications d'histones (par ChIP-seq), d'accessibilité de la chromatine (DNase-seq, FAIRE-seq, ATAC-seq), de méthylation de l'ADN (270K et 450K, RRBS, WGBS) et d'expression des gènes (RNA-seq). L'immense quantité de données générées a ensuite été intégrée grâce à des modèles statistiques sophistiqués : ces résultats montrent que les différents niveaux de régulation épigénétiques s'articulent pour définir **des états chromatinien**s associés à la régulation des gènes (**FIGURE 9A**). Chaque état chromatinien correspond à une combinaison particulière de marques d'histones, à un niveau de méthylation plus ou moins important, et à un degré de compaction de la chromatine. On retrouve ainsi des états chromatinien plutôt actifs, correspondant aux **TSS** et à leurs régions adjacentes (1-3 : H3K4me3 et faible DNAm), au **corps des gènes** (4-5 : H3K36me3 et forte DNAm) et aux **enhancers** (6-7 : H3K4me1, DNAm intermédiaire). Les états inactifs correspondent à **l'hétérochromatine** et à la **chromatine ZNF/éléments répétés** (8-9 : H3K9me3 et forte DNAm), à la **chromatine réprimée par le complexe Polycomb** (13-14 : H3K27me3 et DNAm intermédiaire), et aux domaines de **chromatine bivalente** (10-12), pour laquelle on observe à la fois les marques actives H3K4me1 ou H3K4me3 et la marque répressive H3K27me3, associées à un niveau de méthylation de l'ADN faible. On retrouve cette chromatine bivalente au niveau des régulateurs des gènes importants pour le développement, des gènes soumis à empreinte parentale, et dans les cellules souches embryonnaires : la présence de marques activatrices et répressives permet de moduler rapidement leur expression au cours du développement (Kanayama et al. 2019). Enfin, la **chromatine quiescente inactive**, composant majoritaire du génome, est fortement méthylée mais ne présente pas de modifications d'histones particulières.

Ces différents niveaux de régulation épigénétiques sont influencés par la séquence primaire de l'ADN. En effet, des études de ChIP-seq dans des cellules souches embryonnaires et dans des cellules différenciées ont révélées des mécanismes épigénétiques différents pour la régulation des promoteurs riches en CpG (HCP) et pauvre en CpG (LCP) (**FIGURE 9B**). Les **promoteurs HCP semblent être actifs par défaut**, et requièrent l'addition de multiples marques répressives (DNAm, H3K27me3) pour inhiber leur expression. Par contraste, **les promoteurs LCP semblent être inactifs par défaut** et seraient sélectivement activés, par exemple par des facteurs de transcription (Zhou et al. 2011).

Ainsi, l'observation de la structure chromatinienne et des modifications chimiques de l'ADN ne laisse aucun doute sur le caractère informatif des marques épigénétiques : leur combinaison est étroitement associée à la définition des éléments fonctionnels du génome. Cependant, un message, tout informatif qu'il soit, nécessite d'être encodé puis décodé, interprété et propagé pour être efficace. Dans la partie suivante, nous allons donc présenter les grands mécanismes permettant la mise en place, l'interprétation et la maintenance à travers les générations cellulaires des marques épigénétiques.

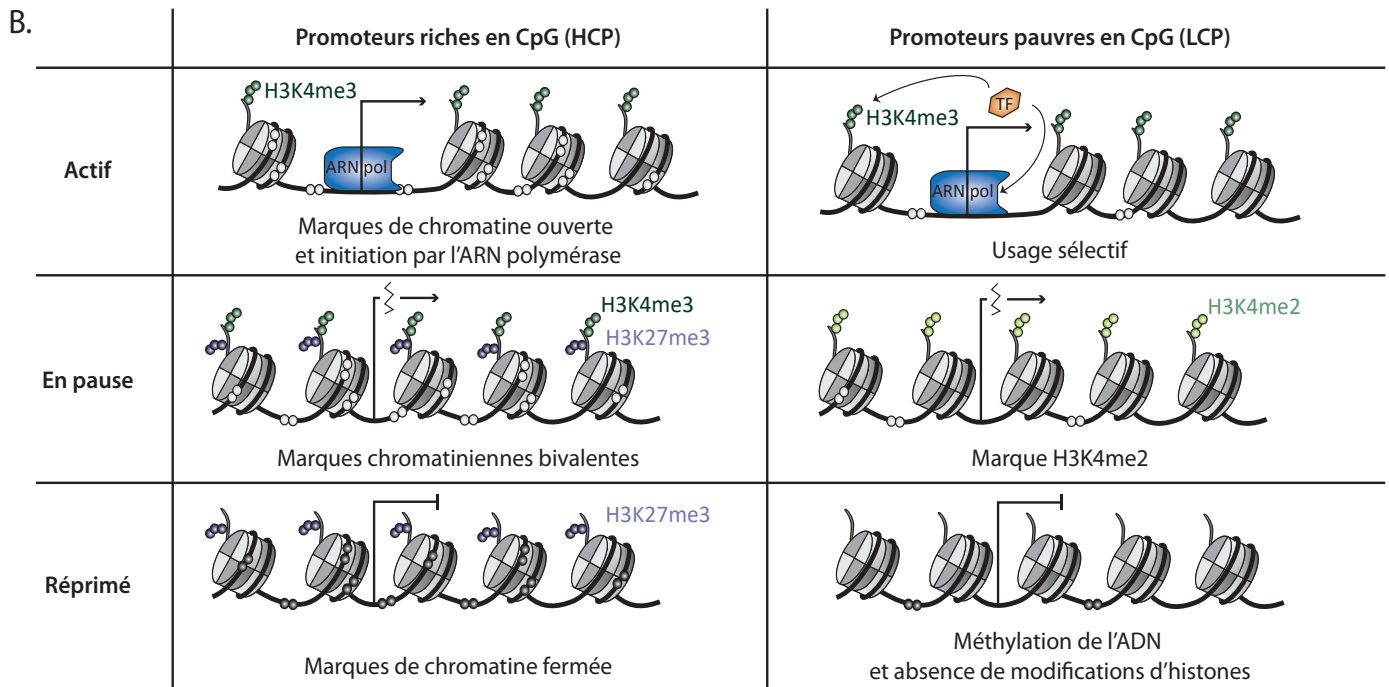
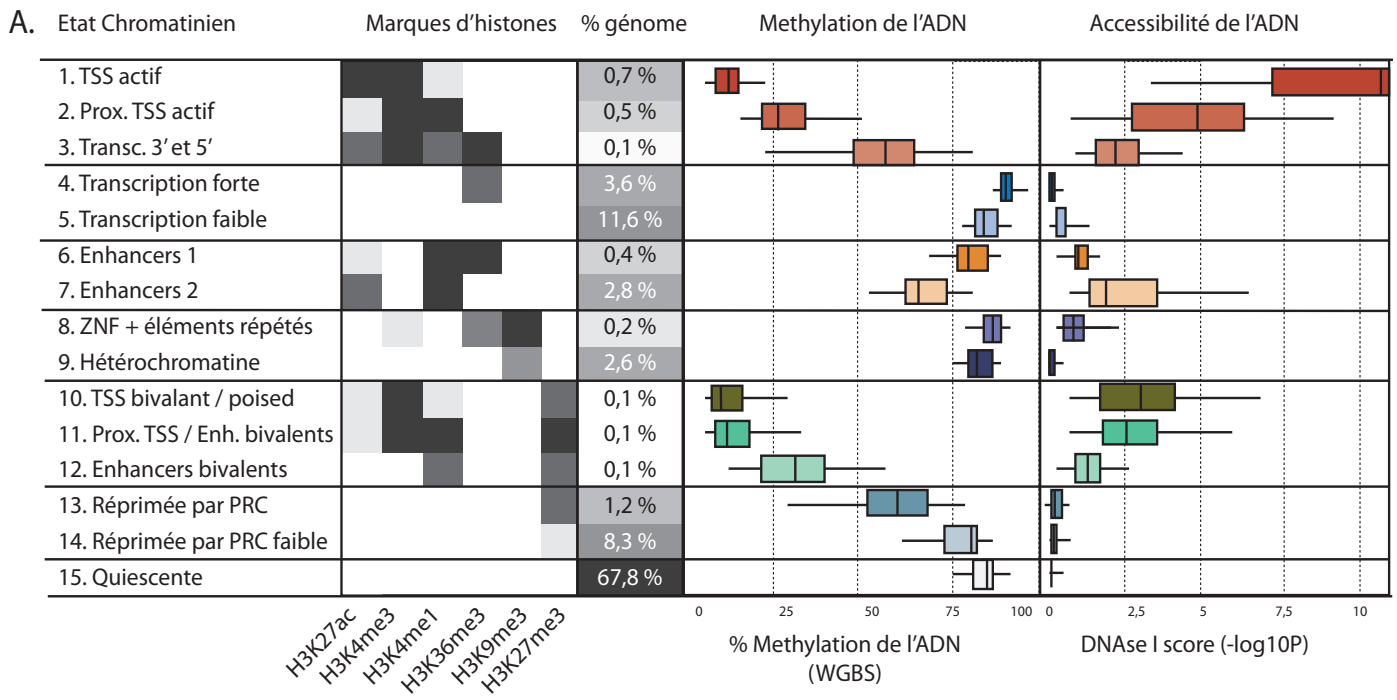


Figure 9 : Combinaisons des différents niveaux de régulation épigénétique

A. La cartographie de la chromatine définit différents états chromatinien ; certains sont plutôt actifs (TSS et leurs régions adjacentes, corps des gènes transcrits, enhancers), les autres sont plutôt inactifs. La quantification de chaque domaine chromatinien montre que le génome est majoritairement constitué de chromatine inactive quiescente. (Adaptée de Kundaje 2015).

B. Il existe un couplage fort entre méthylation de l'ADN et modifications d'histones, qui s'articule à la densité en CpG de la séquence primaire des promoteurs. Les promoteurs riches en CpG font l'objet de modifications chromatinienne distinctes des promoteurs pauvres en CpG, pour un même niveau d'activité transcriptionnelle. (Adaptée de Zhou, Goren & Bernstein 2011)

Partie II

**Comment les marques épigénétiques
sont-elles mises en place, propagées,
interprétées et éventuellement remodelées ?**

1. Lire, écrire et éditer les épigénomes : fonctions des Readers, Writers et Erasers

Les paysages épigénétiques que nous avons décrits sont mis en place, interprétés et entretenus par différentes protéines et ARNs, qui forment une machinerie complexe classiquement séparée en trois catégories (**FIGURE 10**) : les enzymes écrivaines dites « **writers** », les interprètes dits « **readers** », et enfin les enzymes qui enlèvent les marques épigénétiques, dites « **erasers** ».

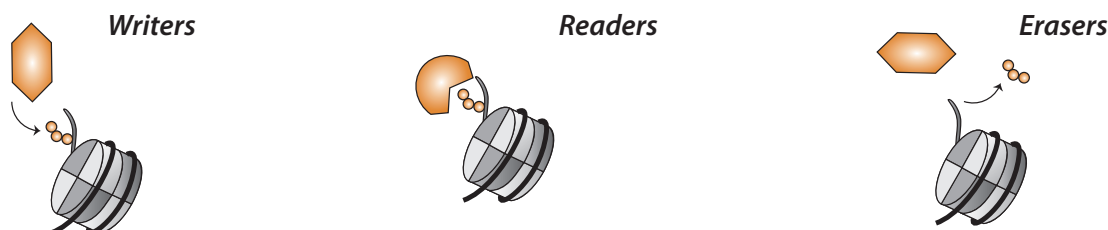


Figure 10 : Writers, Readers & Erasers

Trois fonctions essentielles de la machinerie épigénétique

Les writers sont les enzymes catalysant la mise en place des marques épigénétiques. Ce système présente un certain niveau de spécialisation, qui repose pour partie sur les spécificités de substrat des enzymes : par exemple, en ce qui concerne la méthylation de l'ADN, on connaît chez l'homme 3 ADN-méthyltransférases. DNMT1 a une affinité plus élevée pour l'ADN hémi-méthylé et sera principalement responsable de la maintenance de la méthylation de l'ADN après la réplication semi-conservative. DNMT3A et DNMT3B, elles, ont une plus grande affinité pour les CpG non-méthylés, et seront responsables de l'établissement *de novo* de la méthylation de l'ADN.

Les readers sont des protéines dotées de domaines de reconnaissance pour les sites épigénétiquement modifiés. Les spécificités des différents readers sont également déterminées par la nature de leur domaine de reconnaissance : pour la méthylation des histones, en fonction de la taille et de la nature des acides aminés composant le site de liaison, certains domaines pourront interagir avec les formes mono-, di- ou tri-méthylés des histones. Par exemple, les domaines *chromo-* présentent une poche hydrophobique assez profonde et vont pouvoir se lier aux modifications d'histones tri- ou di-méthylés (H3K9me3, H3K9me2, H3K27me3, H3K27me2), tandis que les domaines MBD, plus petits, interagiront avec des résidus moins encombrants, mono- ou di-méthylés (H3Kme1, H3Kme2, H4Kme1, H4Kme2) (Musselman et al. 2012). Ces readers sont absolument essentiels à la fonction des marques épigénétiques : sans eux, ces modifications n'auraient pas d'effet régulateur sur l'expression des gènes. Cela a été largement démontré pour la méthylation de l'ADN, qui en elle-même n'entraîne pas *in vitro* de baisse directe de l'activité de l'ARN polymérase (Love et al. 1984). La répression transcriptionnelle associée à la méthylation des promoteurs des gènes passe donc par des mécanismes indirects, nécessitant des facteurs dotés de domaines readers capables d'interpréter et de traduire la fonction de ces modifications (Li et al. 1992 ; Okano et al. 1999).

Les *erasers* assurent la réversibilité des marques épigénétiques, et donc la plasticité de ce code. En effet, les marques épigénétiques peuvent être effacées de façon active (par les *erasers* telles que les enzymes *ten-eleven translocation* (TETs) dioxygénases qui catalysent l'hydroxylation puis l'oxydation de la 5meC), mais également de façon passive (par hydrolyse spontanée) et de façon indirecte, telle que la réparation des cytosines déaminées par le système AID/APOBEC qui place une cytosine non méthylée.

Ce système montre un degré important de diversité voire de redondance au niveau génétique : la famille des histones methyltransferases et déméthylases compte facilement plus de 100 gènes au sein du génome des mammifères (Rotili et al. 2011, Copeland et al. 2009), et les lysines méthylées par exemple peuvent être la cible de 50 à 100 protéines *readers* différentes, grâce à leurs domaines « *chromo* », « *Tudor* », « *MBT* », ou « *PHD finger* ». Certains exemples sont répertoriés dans la **FIGURE 11**.






























Méthylation de l'ADN		
 <p>DNA methyl-transférases : DNMT1, DBMT3a, DNMT3b</p>	 CxxC  Methyl Binding Domain  Doigts de zinc  SET et RING finger associated domain (SRA) ... Autres ?	 <p>ten-eleven translocation (TET) methylcytosine dioxygenases : TET1, TET2, TET3</p>
Acétylation des histones		
 <p>Histones Acetyl-transférases (HAT) ex: TAK1</p>	 Bromo  Plant Homeo Domain	 <p>Histones Desacetylases (HDAC) ex: SIRT6 HDAC1...</p>
Méthylation des histones		
 <p>Histones Methyl-transférases (HMT) ex: SET1A/B, MLL1-5 (H₃K₄) SUV39H1/2, G9a, GLP, SETDB1 (H₃K₉) EZH2 (H₃K₂₇) SMYD2 (H₃K₃₆) DOT1 (H₃K₇₉)</p>	 Chromodomain  Plant Homeo Domain (PHD)  Tudor  Malignant Brain Tumor (MBT)  Zinc Finger with conserved Cys and Trp residues (ZN-CW)  Conserved Pro-Trp-Trp-Pro motif (PWWP)  ATRX-DNMT3-DNMT3L (ADD)  Répétition Ankrines  Bromo-Associated Homology (BAH)	 <p>Histones Déméthylases (HDMT) ex: LSD1 (H₃K₄) JHDM2a/b (H₃K₉) JHDM1a/b (H₃K₃₆)</p>
Phosphorylation des histones		
 <p>Kinases ex : MSK1/2 (H₃S₁₂₈)</p>	 BRCA1 C-Terminal (BRCT)  14-3-3  Baculovirus IAP repeats (BIR)	 <p>Phosphatases</p>
Ubiquitylation des histones		
 <p>Ubiquitine ligases ex : Ring1a (H₂AK₁₁₉)</p>	 Ubiquitine Binding Domain (BUD)	 <p>Desubiquitinases (DUB)</p>

Figure 11 : Writers, Readers & Erasers

Enzymes et domaines de reconnaissance associés aux principales marques épigénétiques. Données tirées de Bannister & Kouzarides 2011; Rothbart & Strahl 2014

2. Coordination épigénétique grâce à l'association des acteurs chromatinien

Les résultats du projet ENCODE notamment montrent que la méthylation de l'ADN et les modifications des histones sont des processus interdépendants. Prenons un exemple concret : pour réprimer un domaine chromatinien, nous pouvons imaginer plusieurs séquences (FIGURE 12). Dans le premier modèle, la méthylation de l'ADN entraîne les modifications des histones : les CpG sont méthylés *de novo* par les DNMT3A et DNMT3B. Les CpG méthylés entraîneraient le recrutement de complexes associant des MBD et des HDACs, tels que le complexe MeCP2 – Sin3a – HDAC, qui pourront alors désacétyler les histones et renforcer la répression des gènes. Cette conformation chromatinienne attire ensuite le recrutement d'HMT telles que Suv39h ou G9a, méthylant les résidus H3K9 et stabilisant l'inactivation chromatinienne. Dans le second modèle, ce sont les modifications des histones qui entraînent la méthylation de l'ADN. Les marques H3K9me3 sont reconnues par le facteur HP1, qui recrute les enzymes DNMTs sur cette région hétérochromatinisée, assurant alors la méthylation des CpG. Enfin, dans un troisième modèle, les remodeleurs chromatinien recrutent et facilitent l'accès aux différentes enzymes épigénétiques, permettant la stabilisation des états de compaction et de décompaction.

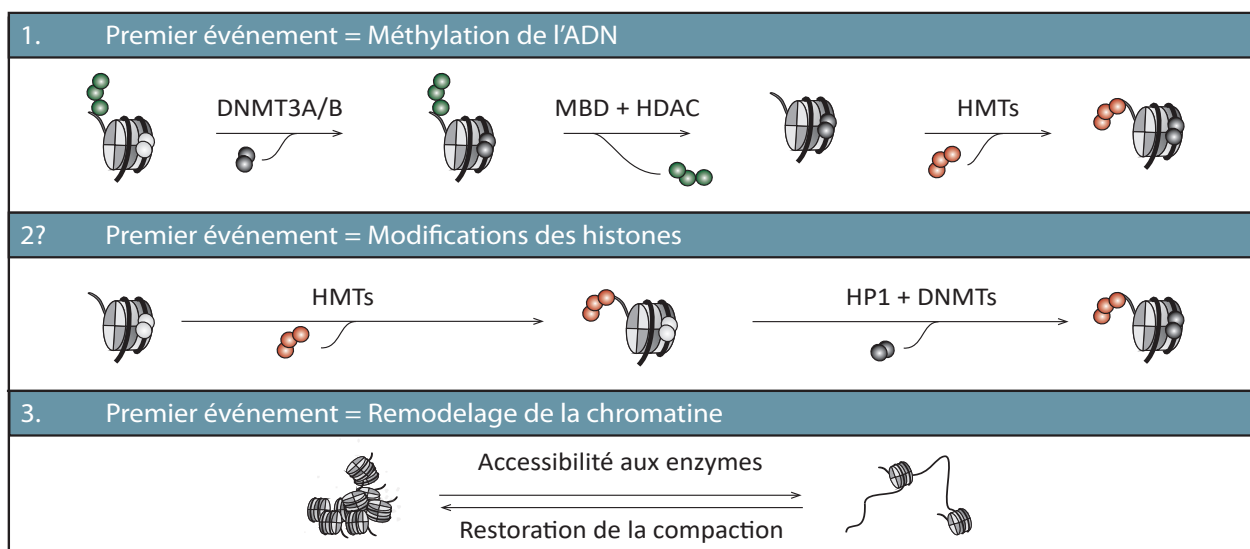


Figure 12 : Coopération des acteurs épigénétiques lors de répression transcriptionnelle

Pour qu'une telle coopération soit possible, les acteurs épigénétiques combinent plusieurs domaines protéiques : un domaine catalytique (*writer/eraser*), un voire plusieurs domaines de reconnaissance des marques épigénétiques (*readers*), et des domaines d'interactions (FIGURE 13A). Certains mécanismes ont été extensivement décortiqués : par exemple, dans le cas de l'exclusion entre méthylation de l'ADN et H3K4me4 (promoteurs actifs), la sélectivité des DNMT3A et DNMT3B pour les promoteurs inactifs est assurée par un mécanisme d'inhibition intra-moléculaire par différents domaines *readers* (Rajavelu 2018, Guo 2015). Cependant, tous les modulateurs épigénétiques ne possèdent pas ces différents modules *readers* : bien souvent, **la coopération des modulateurs épigénétiques coordonnent l'association des marques épigénétiques**, en formant de véritables complexes qui

stimulent le recrutement ou au contraire l'exclusion de certains partenaires. C'est le cas de DNMT1, une protéine d'environ 1600 acides aminés possédant plusieurs domaines non-catalytiques, et qui joue le rôle de véritable plate-forme d'assemblage pour les acteurs impliqués dans les fonctions chromatinienne (FIGURE 13B) (pour revue voir Laisné et al. 2018). Les enzymes *writers* ou *erasers* des marques épigénétiques sont donc, grâce à ces domaines d'interactions protéine-protéine ou protéine-ARN, des acteurs centraux de complexes multiprotéiques qui articulent les activités de modifications épigénétiques aux fonctions chromatinienne que sont la réplication, la réparation et la régulation de l'expression des gènes.

Les mécanismes épigénétiques assurent la mise en place d'une information qui perdure dans le temps : trois types d'événements cellulaires sont donc particulièrement intéressants du point de vue de l'épigénétique. Il s'agit de la **différenciation cellulaire**, où les mécanismes épigénétiques mettent en place *de novo* la configuration chromatinienne typique d'un certain lignage cellulaire ; de la **division cellulaire**, où la réplication de la chromatine implique la restauration des modifications épigénétiques sur les éléments néo-synthétisés ; et enfin la **réponse et l'intégration de signaux**, qui impliquent parfois la mémoire épigénétique afin de faire perdurer ces signaux dans le temps.

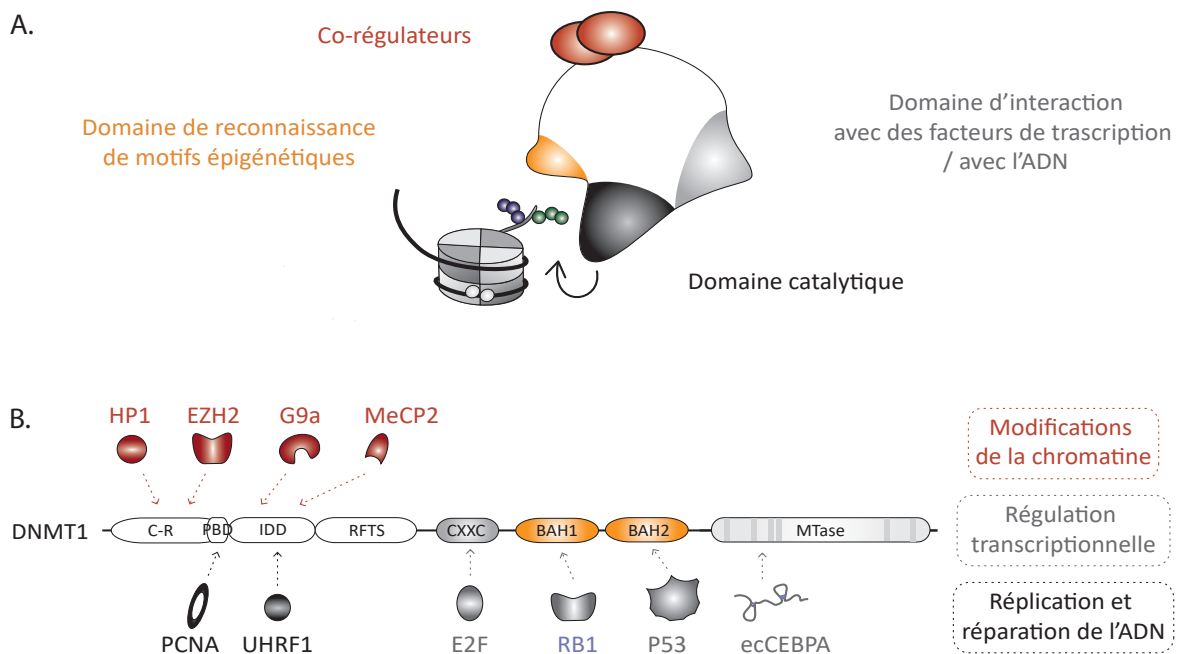


Figure 13 : Coopération des fonctions chromatinienne

A. Exemple d'association de domaines sur un modulateur épigénétique

C. DNMT1, une plateforme d'assemblage pour assurer les fonctions chromatinienne ? (Adapté de Laisné et al. 2018)

3. Mise en place de novo des marques épigénétiques lors du développement

Les premières étapes de différenciation cellulaire ont été extensivement décrites chez la souris : il s'agit de la différenciation entre le trophoctoderme, puis l'ectoderme primitif et l'endoderme primitif lors du stage blastocyte (**FIGURE 14**). De ces trois tissus embryonnaires se différencieront toutes les lignées cellulaires de l'organisme, dont deux compartiments en particulier se définiront très tôt : les cellules germinales, totipotentes, qui maintiennent un état de totipotence épigénétique au cours du temps ; et les cellules somatiques. **Les gènes spécifiques de la lignée germinale**, encodant des fonctions importantes pour la formation des gamètes comme la méiose, **vont donc être verrouillés très précocement dans les cellules somatiques**, où ils doivent rester solidement réprimés (*Borgel et al. 2010*). Lors de ces événements, la différenciation peut être vue comme un mécanisme limitant la plasticité cellulaire : au niveau épigénétique, elle consiste à **poser des barrières inhibant les possibilités de reprogrammation cellulaire, et verrouillant le destin des cellules** (*Smith, Sindhu & Meissner 2016*). De fait, on observe rapidement une diminution de l'accessibilité de la chromatine au cours des premiers événements de différenciation de la gastrulation, avec un gain global et progressif de méthylation de l'ADN, une diminution progressive du niveau de H3K27me3 et un gain, progressif également, des marques activatrices H3K4me3 (**FIGURE 14**).

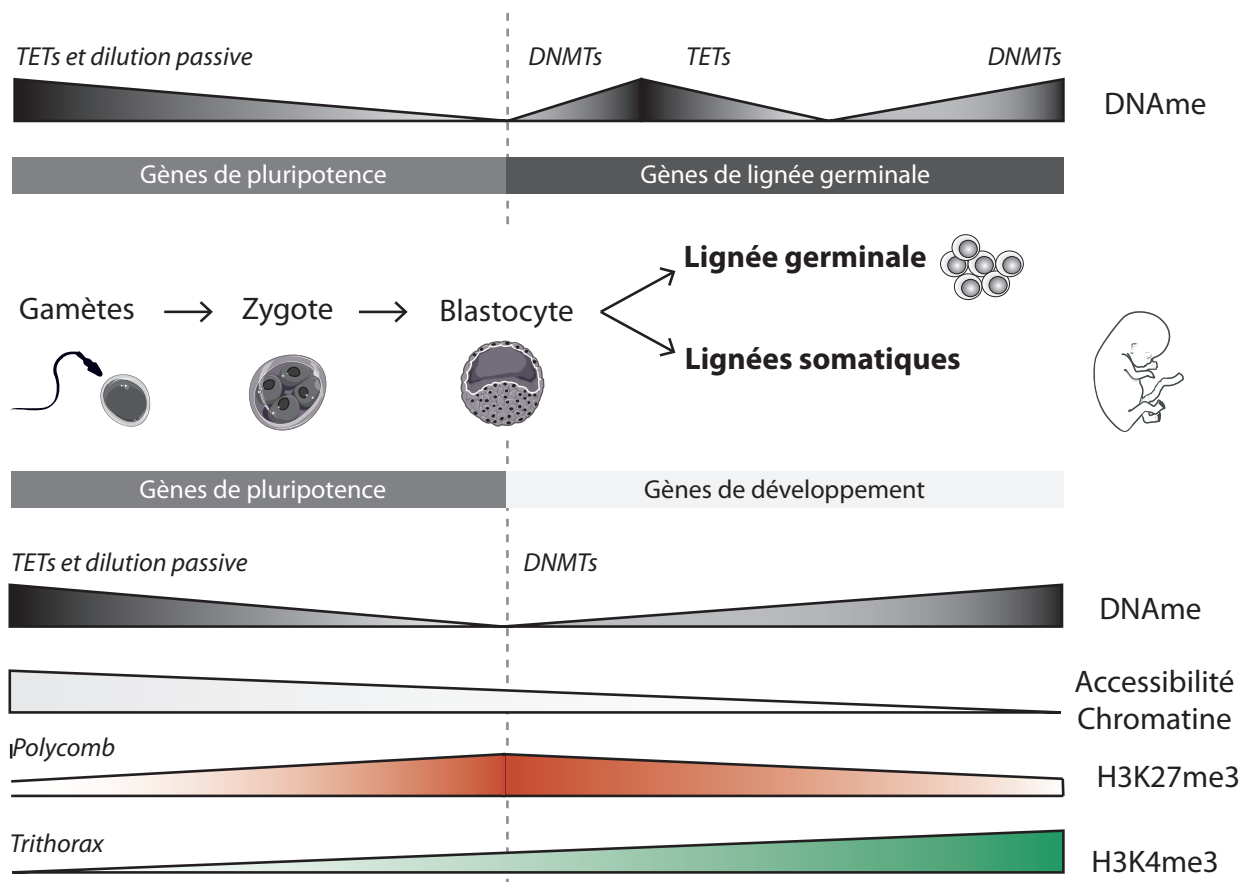


Figure 14 : Développement embryonnaire et différenciation cellulaire

Au cours du développement embryonnaire, le programme transcriptionnel va bifurquer lors de la transition mid-blastocyttaire : pour les cellules somatiques, de l'expression de gène de pluripotence vers l'expression de gènes de développement ; pour les cellules germinales vers l'expression de leurs gènes de lignée. Ce changement de programme est assuré par l'établissement des épigénomes spécifiques.

Ces modifications épigénétiques sont le résultat de l'activité de deux groupes de protéines, initialement décrites chez la drosophile : **les protéines du groupe *Polycomb*** (PcG : le complexe PRC2, pour H3K27me3, et PRC1, pour H2K119Ub) **et du groupe *thritorax*** (TrxG : les HMTs MLL1, MLL2 et hSET1, pour H3K4me3). Soulignant le rôle crucial de ces protéines dans la différenciation cellulaire, les altérations des complexes PcG et TrxG sont associées à de multiples tumeurs humaines (pour revue : [Pasini & Di Croce 2016](#)).

Ces premières modifications épigénétiques entraîneront le recrutement de différents facteurs de transcription, selon les patrons propres à chaque lignage (FIGURE 15), qui à leur tour recruteront d'autres modulateurs épigénétiques afin de verrouiller plus solidement les états transcriptionnels, la méthylation de l'ADN étant souvent le dernier événement. Dans ce sens, l'analyse de 38 facteurs de transcription dans les trois tissus embryonnaires montre la présence de sites de liaisons spécifiques à chaque tissu ([Tsankov et al. 2015](#)), ce qui suppose une différenciation précoce des régions de chromatine ouvertes ou fermées. Pour complexifier un peu le tableau, cette étude a montré que la reconnaissance d'un même site par un même facteur A peut entraîner dans certains tissus la déméthylation de ce site, ou son hyperméthylation dans un autre tissu : ces résultats montrent combien l'interprétation épigénétique de la fixation d'un facteur de transcription dépend du contexte cellulaire ([Tsankov et al. 2015](#), [Kim et al. 2020](#)).

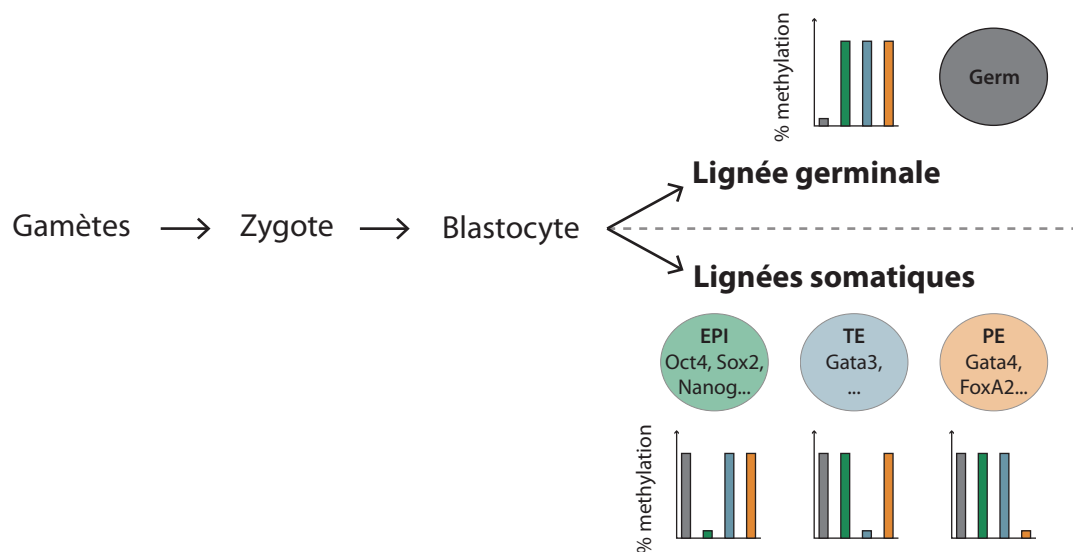


Figure 15 : Divergence précoce des lignées germinales et somatiques

Certains facteurs de transcription spécifiques de chaque tissu embryonnaires sont spécifiés. EPI : épiblaste ; TE : trophectoderme ; PE : épiderme primitif

Au cours du développement embryonnaire, la mise en place des modifications épigénétiques précède les modifications transcriptionnelles : cela suppose que les premiers modulateurs épigénétiques devront reconnaître leurs séquences cibles dans un contexte chromatinien non permissif à la transcription. Les **facteurs pionniers** sont de ceux-là : ils sont capables d'interagir avec l'ADN nucléosomal, avant l'activation des leurs *enhancers* cibles, et d'entraîner le recrutement de modulateurs épigénétiques afin de remodeler localement l'environnement chromatinien et de permettre aux autres facteurs de transcription d'accéder à leurs motifs de reconnaissance sur la séquence nucléotidique.

L'exemple le plus extrême est celui de la reprogrammation directe de cellules différenciées en cellules pluripotentes induites (iPSCs) grâce à l'expression ectopique de 4 facteurs de transcription (Takahashi & Yamanaka 2016) : Oct4, Sox2, Klf4 et Myc suffisent à déstabiliser les réseaux transcriptionnels en place et à induire des changements majeurs de la structure chromatinienne, permettant d'établir un paysage épigénétique similaire à celui des cellules souches (pour revue : Nashun, Hill & Hajkova 2015). Les facteurs Oct4, Sox2 et Klf4 sont effectivement des facteurs pionniers, ainsi que les facteurs de la famille FoxA ; la dérégulation de tels facteurs peut jouer un rôle important dans le développement de certaines pathologies humaines.

4. Maintien des marques épigénétiques à travers les divisions cellulaires

Juste après la division cellulaire, on observe une diminution drastique de la quantité globale d'ADN méthylé, ainsi que des marques d'histones H3K27me3, H3K9me3 et H3K4me3 (FIGURE 16), soulignant la nécessité de restaurer ces marques épigénétiques après la duplication du matériel génétique. Elles le seront progressivement, au cours du cycle cellulaire : la réplication de l'ADN implique la réplication globale de la structure chromatinienne, mais ces deux événements se font en quelque sorte concurrence. En effet, répliquer la molécule d'ADN suppose de déplacer, puis de replacer les nucléosomes afin de ménager un accès à la machinerie de réplication. Ces déplacements sont effectués par des chaperons d'histones et doivent restaurer à l'identique les marques chromatiniennees sur les deux molécules filles. Cela exige d'une part la néosynthèse d'histones pour doubler la quantité de nucléosomes, d'autre part la propagation des modifications post-traductionnelles existantes sur les histones néosynthétisées (FIGURE 17) : d'où ce délai observé pour la restauration des niveaux globaux de modifications d'histones (voir pour revue Stewart-Morgan et al. 2020). De même, la méthylation des CpG doit elle aussi être reproduite symétriquement sur le brin d'ADN néosynthétisé (FIGURE 17).

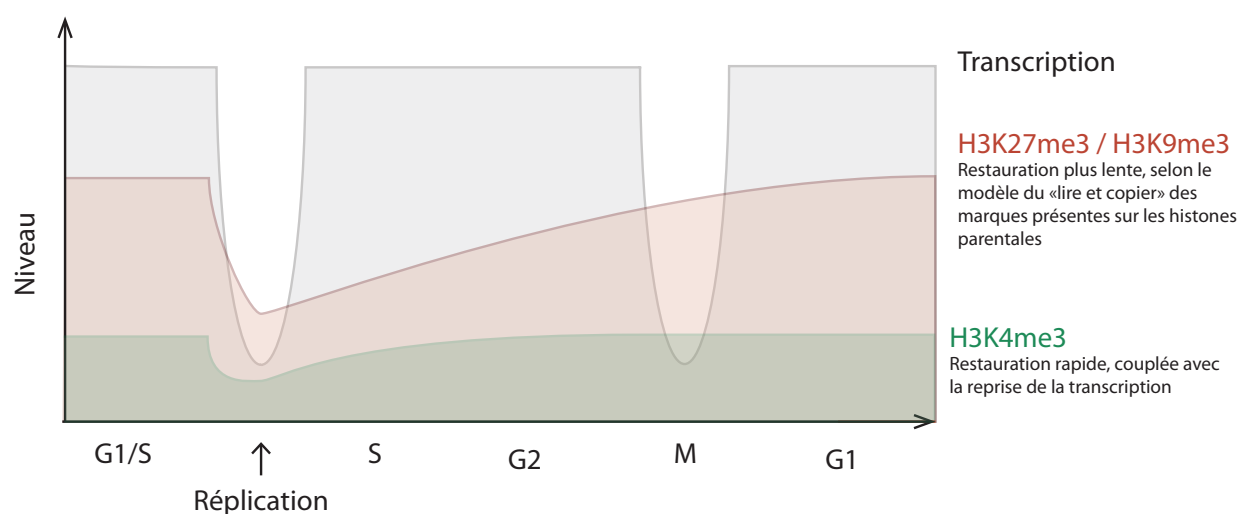


Figure 16 : Cinétique de la restauration des modifications d'histones au cours du cycle cellulaire

(Adapté de Stewart-Morgan et al. 2020)

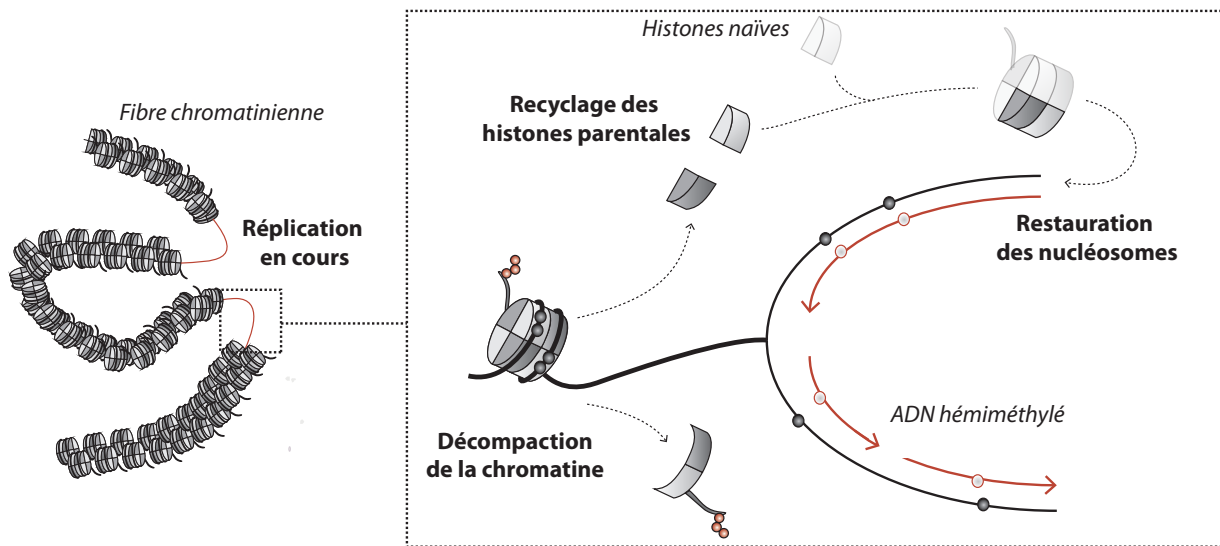


Figure 17 : Division cellulaire et réplication de la chromatine

Représentation des foyers de réplication (en rouge) et de la fourche de réplication, avec les principaux événements chromatiniens. Après le passage de la fourche de réplication, il faut restaurer la méthylation de l'ADN sur le brin néosynthétisé et les modifications post-traductionnelles des histones naïves.

Plusieurs modèles peuvent être proposés pour rendre compte des mécanismes de maintenance des épigénomes. Tout d'abord, en ce qui concerne la méthylation des CpG, le modèle le plus simple **exploite les propriétés semi-conservatives de la réplication pour copier le patron de méthylation du brin parental sur le brin néosynthétisé** (Holliday & Pugh 1975, Riggs et al. 1975), grâce à l'affinité préférentielle de la méthyltransférase de maintenance DNMT1 pour les CpG hémi-méthylés (Pradhan et al. 1999). Cependant la maintenance de la méthylation de l'ADN recèle encore certains mystères et ce modèle se heurte à des observations contradictoires : des expériences de méthylation *in vitro* montrent que la fidélité de ce processus au cours de la réplication est assez faible. En réalité, dans ce **modèle on estime à 5% le nombre de CpG qui échappent à la méthylation par division cellulaire** (Wigler et al. 1981). Ce taux d'erreur est retrouvé dans des lignées cellulaires clonales (Riggs et al. 1998) : si on ne compte que sur la méthylation associée à la réplication, on s'attendrait à observer une dilution rapide des 5mC au fil des divisions cellulaires. De plus, non contentes d'échouer à reproduire totalement les patrons de méthylation au cours de la réplication, des expériences ont montré que de nombreux CpG étaient incorrectement méthylés *de novo* au cours de la division cellulaire. Cependant, et malgré ces erreurs, on observe une conservation remarquable des patrons de méthylation au cours du développement : des mécanismes alternatifs ou complémentaires à l'action de DNMT1 agissent pour assurer la propagation des patrons de méthylation d'une génération cellulaire à l'autre (Rhee et al. 2000, Jaenish et al. 1997). La méthylation de l'ADN est particulièrement stable au niveau des îlots CpG (Pfeifer et al. 1990) : il semble qu'en matière de méthylation de l'ADN, la densité en CpG soit un élément important pour assurer la robustesse de l'état de méthylation (ou de non méthylation).

Au cours de la réplication, la coopération entre les modulateurs épigénétiques et la complémentarité de leurs domaines de reconnaissances assurent une fois encore le couplage de la méthylation de l'ADN et des différentes modifications d'histones : ainsi l'association d'UHRF1 avec la marque répressive H3K9me3 favorise le recrutement de DNMT1 sur ces sites et facilite la maintenance des patrons de méthylation de l'ADN (Rothbart et al. 2012). C'est pourquoi les dérégulations de ces partenaires peuvent également être associées à des pathologies humaines sévères : par exemple, dans les cancers du sein, la répression épigénétique du gène suppresseur de tumeur BRCA1 est notamment associée à la surexpression de UHRF1 (Jin et al. 2010).

5. Propagation épigénétique des signaux dans le temps

Selon une vision classique de la génétique, l'expression des gènes dépend de la disponibilité d'une combinaison de facteurs de transcription et de leur association avec les éléments de contrôle du gène régulé. Cependant, la liaison des facteurs de transcription est bien souvent transitoire et est perdue au cours de la division cellulaire. **Pour que les patrons d'expression génétiques perdurent dans le temps, le recrutement spatio-temporel des différents facteurs de transcription doit être reproduit à chaque division cellulaire.** C'est pourquoi l'accessibilité de la chromatine, condition sine qua non du recrutement des facteurs de transcription non-pionniers, encode cette information dans la conformation 3D héréditaire de la chromatine, et assure la propagation de signaux comme l'état d'activation d'un promoteur ou la définition des centromères à travers les générations cellulaires (FIGURE 18). L'identité des cellules, acquise lors des processus de différenciation cellulaire, est ainsi stabilisée par cette mémoire épigénétique au fil des divisions.

La mémoire épigénétique permet également de propager des signaux à moyen et court terme, en répondant à des stimuli externes ou internes (Jaenisch & Bird 2003). On peut penser à la réponse aux dommages à l'ADN, où la réparation de la séquence nucléotidique entraîne un remaniement de la chromatine important, avec notamment l'incorporation d'histones naïves et la synthèse d'un ADN hémiméthylé. Les mécanismes de restauration des marques épigénétiques sont parfois incomplets, laissant alors une cicatrice à l'endroit de la réparation qui pourrait contribuer à la mémoire cellulaire de la zone de dommage (pour revue : Dabin, Fortuny & Polo 2017).

L'intégration de signaux primaires venant des voies de signalisation cellulaire peut également passer par des modifications épigénétiques, qui propageront ce signal dans le temps. **Peu de choses sont encore établies concernant ces interconnexions entre signalisation cellulaire et modulateurs épigénétiques : j'ai pu explorer l'une de ces relations dans un article publié en 2020,** démontrant une connexion entre la voie non-canonique du TGF-beta, les enzymes épigénétiques SIRT6 et KAT2A, et le gène *ADAM12* soumis à des régulations épigénétiques (Naciri et al. 2020). Enfin, toujours dans l'idée de traduire épigénétiquement des signaux cellulaires, il ne faut pas oublier que de nombreuses enzymes de modifications de la chromatine requièrent des cofacteurs, tels que l'ATP pour les kinases, l'acetyl-CoA pour les HATs et SAM pour les HKMT / DNMTs. Or les concentrations de ces cofacteurs

reflètent directement les variations environnementales (comme le régime alimentaire par exemple : voir [Guarante & Picard 2005](#)). Ainsi, des conditions métaboliques particulières sont susceptibles d'être traduites épigénétiquement, influençant la régulation de certains gènes à plus ou moins court terme : un exemple extrêmement visuel de ce phénomène est donné par les variations de la couleur du pelage du modèle murin *agouti viable yellow*, directement lié à la disponibilité en donneurs de méthyle durant la période gestationnelle et donc à la qualité du régime alimentaire de la mère ([Wolff et al.1998](#), voir Introduction).

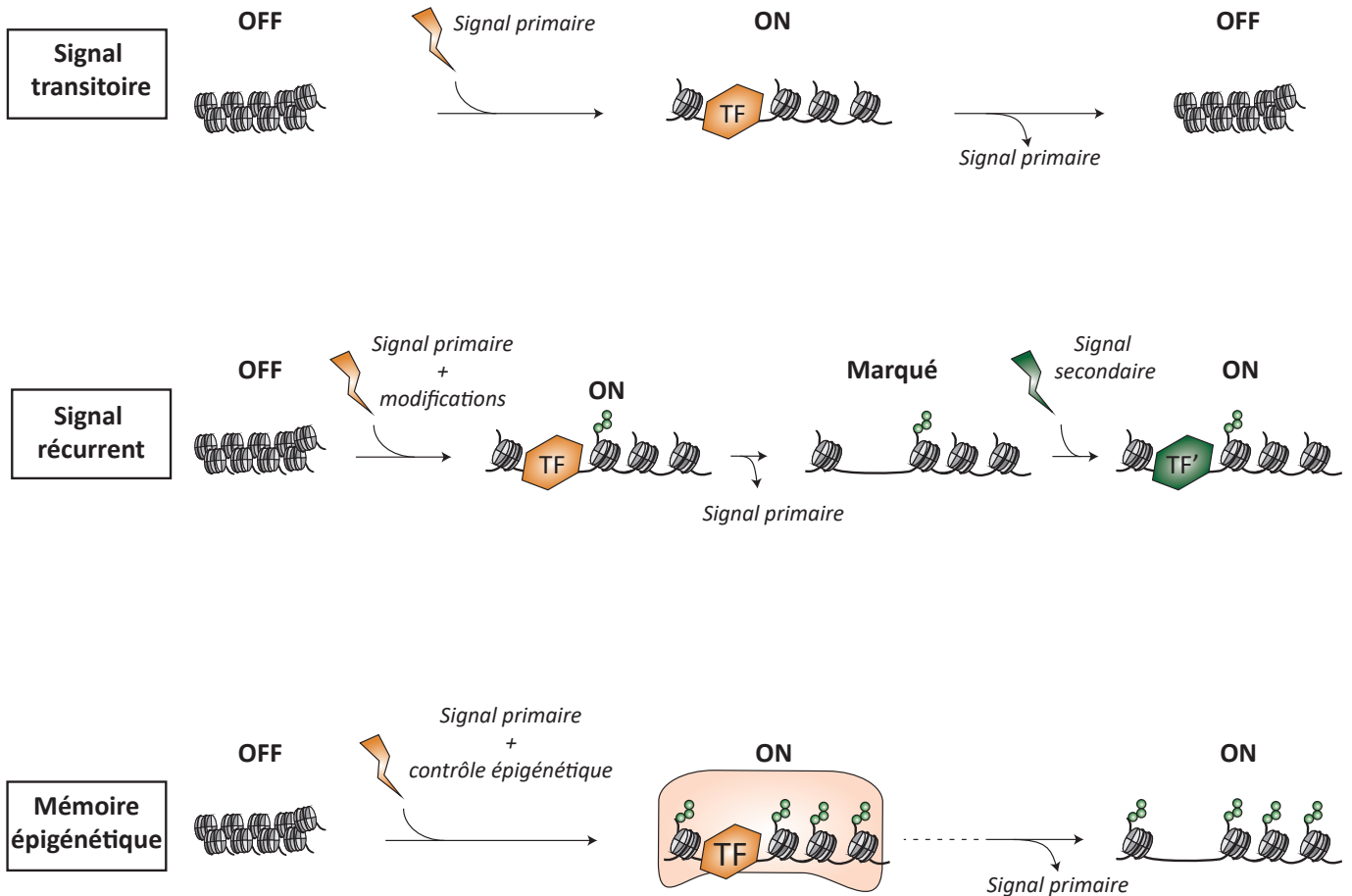


Figure 18 : Potentialisation épigénétique des signaux cellulaires

(Adapté de Allis DC et al. 2006)

Partie 3

**Quelles sont les fonctions essentielles
des marques épigénétiques ?**

Au cours des paragraphes précédents, nous avons construit un portrait des principales modifications épigénétiques des cellules humaines, auquel se surimposent les acteurs majeurs impliqués dans leur mise en place, de leur maintien au cours des divisions cellulaires et de leur interprétation de ces marques. Cette machinerie épigénétique est au service de fonctions cellulaires essentielles, dont l'orchestration est cruciale pour les cellules saines et dont la dérégulation est fréquemment observée dans les cellules cancéreuses.

1. Restreindre le répertoire d'expression des gènes

Bien qu'elles partagent le même génome, les différents types cellulaires d'un organisme présentent des morphologies variées, jouent des rôles spécifiques et répondent différemment aux signaux environnementaux, développementaux ou métaboliques. Ces caractéristiques définissent l'identité des cellules, et sont traduites notamment au niveau de l'expression des gènes grâce à la participation des épigénomes.

D'une part, les épigénomes mis en place au cours de la différenciation cellulaire suffisent à identifier les différents lignages. Les grands projets de séquençage des épigénomes humains tels que le *NIH Roadmap Epigenomics Consortium* ont généré des collections de données sur plus d'une centaine de cellules primaires et de tissus humains, et l'analyse intégrative de ces données démontre clairement l'existence de profils épigénétiques spécifiques aux différents tissus (ENCODE 2015). Par exemple, dans une étude préalable au projet ENCODE, Ziller et al. ont démontré qu'il était possible de classer robustement différents tissus humains sur la seule base des valeurs de méthylation d'un petit nombre de régions génomiques (FIGURE 19, Ziller et al. 2013). Ces régions particulièrement informatives se trouvent être différentiellement méthylées selon le lignage : de façon intéressante, elles correspondent majoritairement à des régions de type *enhancer* (Ziller et al. 2013), ce qui laisse deviner leur rôle dans la mise en place de transcriptomes spécifiques. Notons au passage que **le tissu testiculaire ségrège** séparément des autres tissus somatiques fœtaux et adultes, soulignant cette divergence précoce au cours du développement entre les épigénomes de la lignée germinale et des lignées somatiques. On retrouve cette même association au niveau des autres marques de la chromatine (ENCODE 2015) : **les marques épigénétiques les plus discriminantes pour identifier les lignages cellulaires humains se trouvent principalement associées aux régions *enhancer*.** L'analyse par *Gene Ontology* des gènes situés dans le voisinage de ces modules épigénétiques révèle leur association avec les fonctions spécifiques de ces tissus : par exemple, les *enhancers* spécifiques des cellules souches présentent un enrichissement en gènes voisins impliqués dans le développement.

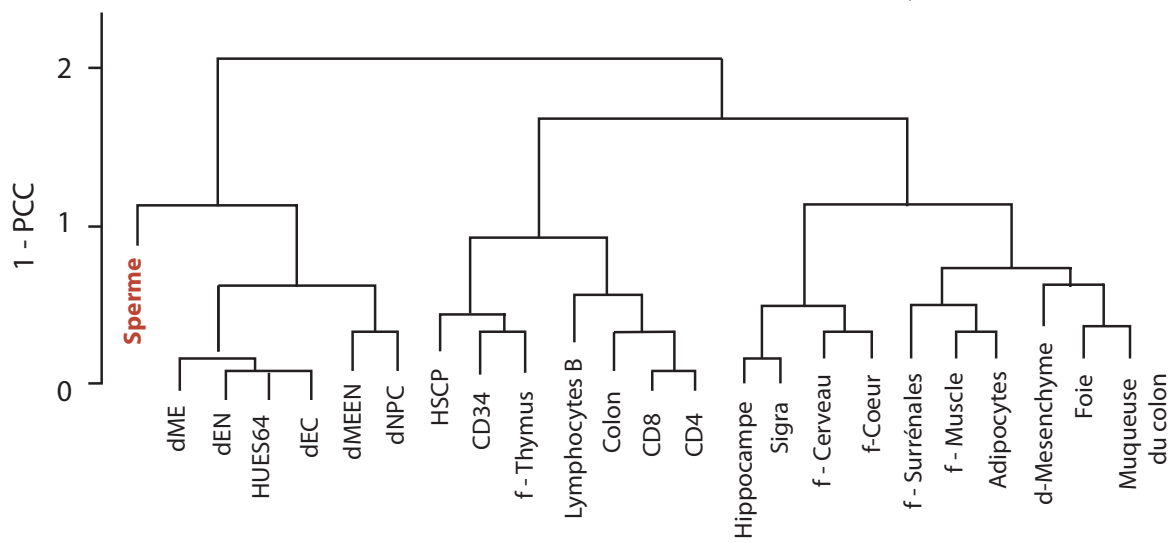


Figure 19 : Classification de tissus humains à partir de données de méthylation d'ADN

Clustering hiérarchique utilisant le coefficient de corrélation de Pearson (PCC) des régions différenciement méthylés d'échantillons humains (Adapté de Ziller et al. 2013).

f: fetal

D'autre part, les différents types cellulaires peuvent être identifiés tout aussi robustement par leurs signatures transcriptomiques. Les projets menés notamment par le consortium *Human Protein Atlas* (HPA - RNAseq, [Uhlen et al. 2015](#)), *FANTOM* (CAGE, [Yu et al. 2015](#)) ou encore le projet *GTEX* (RNAseq, [Keen & Moore 2015](#)) ont utilisé les technologies de séquençage à haut débit pour analyser les transcriptomes humains. Plus d'une vingtaine de tissus et organes humains ont été analysés, permettant de **classer les gènes humains en différentes catégories** : les gènes exprimés préférentiellement dans un tissu particulier (souvent appelés gènes **tissu-spécifiques ou tissu-restreints**), les gènes exprimés dans un groupe de tissus (2 à 7 tissus), ainsi que les **gènes ubiquitaires** (dont les gènes de ménage) exprimés dans tous les tissus (**FIGURE 20A**) (pour revue : [Uhlen et al. 2016](#)). Le terme de gène tissu-spécifique peut prêter à confusion car leur définition dépend fortement des valeurs seuils choisies et de la technique de détection employée (l'influence de ces paramètres a notamment été évaluée dans [Kryuchkova-Mostacci & Robinson-Rechavi 2016](#)), et seuls quelques gènes caractéristiques tels que les gènes codant pour l'insuline ou la PSA ne sont détectés que dans un seul tissu ([Uhlen et al. 2015](#)). Nous retiendrons pourtant ce terme, qui a les qualités de ses défauts : sa simplicité le rend très expressif, et résume aisément l'association préférentiel entre certains gènes et tissus. **Les deux projets GTEX et HPA s'accordent pour définir le tissu testiculaire comme étant le tissu ayant le plus grand nombre de gènes spécifiques (FIGURE 20B).** De ces analyses, on en déduit que les différents types cellulaires expriment un sous-ensemble de gènes tissu-spécifiques, qui leur est propre, des gènes associés à un groupe de tissus et des gènes ubiquitaires. Une analyse fonctionnelle par *Gene Ontology* sur les différents groupes de gènes tissus-spécifiques a été réalisée et les résultats sont cohérents avec les fonctions associées à chaque tissu : par exemple, les gènes spécifiques du foie sont associés aux fonctions de sécrétion de la bile, de détoxification et de processus métaboliques tels que le stockage du glycogène ([Uhlen 2015](#)). **Ces gènes tissu-spécifiques représentent une sorte de carte d'identité des cellules, symétriquement aux modules épigénétiques tissu-spécifiques.**

Deux autres cas importants de contrôle épigénétique de l'activité des gènes concernent l'expression mono-allélique des gènes soumis à empreinte parentale, et de certains gènes présents sur le chromosome X chez les femelles. **L'empreinte parentale** est responsable de la non-équivalence des génomes paternels et maternels. Dans les années 1980, les embryologistes Solter et Surani échouent à produire un embryon à partir de deux génomes d'origine maternelle ou bien de deux génomes d'origine paternelle : dans les deux cas, on observe une létalité embryonnaire précoce ([McGrath & Solter 1984](#) ; [Surani et al 1984](#)). Dans les années 1990, le mécanisme à l'origine de ce phénomène de non-équivalence des génomes maternels et paternel a été identifié : la méthylation de l'ADN y joue un rôle central. Le maintien de l'expression mono-allélique des gènes soumis à empreinte est essentielle à l'homéostasie des cellules ([Holm et al. 2005](#)). **L'inactivation du chromosome X** a été décrit en 1961 par Mary Lyon : elle observe que contrairement aux mâles, les femelles des mammifères peuvent présenter des phénotypes « en mosaïque » de la couleur du pelage. L'explication est la suivante : la couleur du pelage est encodée par différents allèles situés sur les chromosomes X, et les variations résultent de l'inactivation aléatoire d'un des chromosomes X dans chacune des cellules de l'embryon, suivie de la transmission stable de

cet état de répression au cours des divisions cellulaires. Les mécanismes moléculaires à la base de la mise en place et du maintien de l'inactivation du chromosome X ont par la suite été explorés (pour revue : Vallot et al. 2016, Patrat et al. 2020). Notons **qu'un des événements clés de la transformation cancéreuse pourrait être dans certains cas l'échappement à l'inactivation d'oncogènes liés au X**, et que des défauts de l'inactivation d'un des chromosome X pourrait jouer un rôle dans la progression tumorale (Chaligné & Heard 2014 ; Spatz, Bork & Feunteun 2020). Dans le cas des cancers du sein, il a été montré un lien entre la perte du gène suppresseur de tumeur *BRCA1* et la réactivation du X inactivé.

Ces patrons d'expression spécifiques permettent de comprendre pourquoi certaines maladies héréditaires n'affectent que quelques organes particuliers, bien que les mutations incriminées soient présentes dans chaque cellules (pour revue Hekselman & Yeger-Lotem 2020). **Il en est de même pour les mutations oncogéniques** : les porteurs et porteuses de ces altérations présentent un risque élevé de certains types de cancers particuliers, ainsi les mutations germinales de *BRCA1/2* augmentent particulièrement le risque de cancers du sein et de l'ovaire chez les femmes. Ces associations peuvent être dues à une expression exclusive ou préférentielle des gènes mutés dans certains tissus, à la mutation d'un élément crucial à la stabilisation de la structure chromatinienne spécifique à ce tissu, ou encore à l'altération d'un réseau de régulation ou d'un composant d'un complexe protéique spécifique à ce type cellulaire.

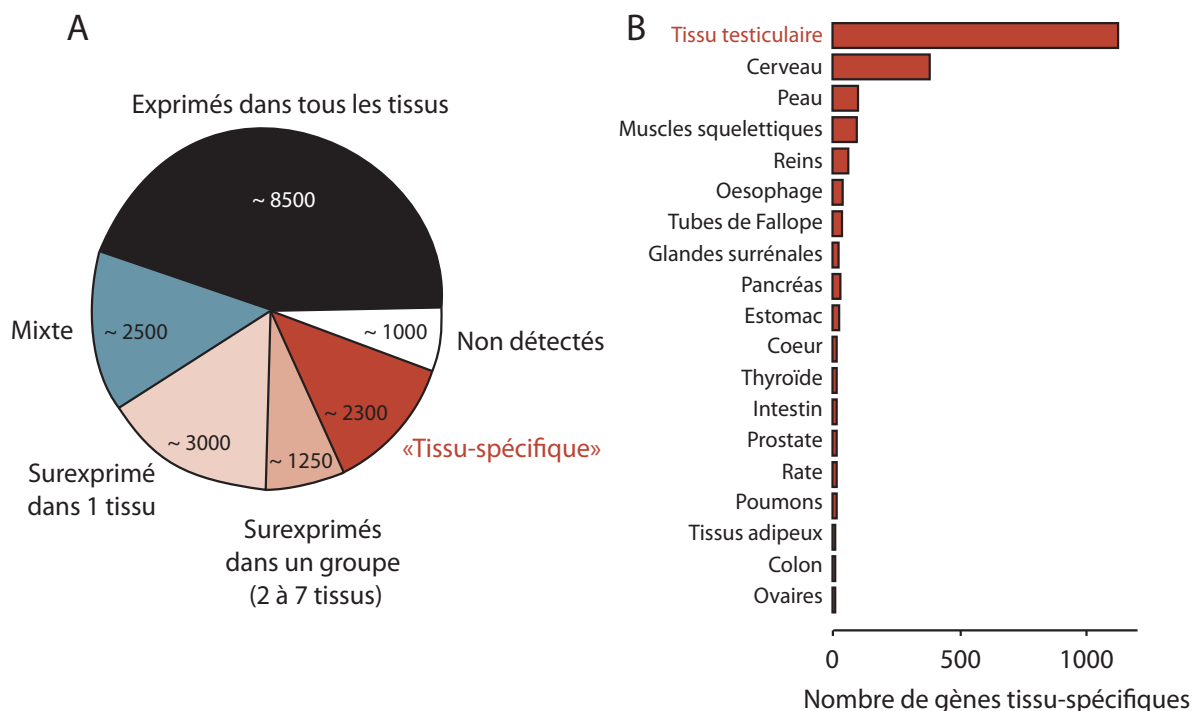


Figure 20 : Classification des gènes codants selon leur spécificité d'expression

Ces graphiques ont été générés à partir des données des projets HPA et GTEx, reprenant les résultats de leurs analyses. Les critères utilisés pour définir ces familles de gènes sont les mêmes dans les deux projets (seuil de détection fixé à 0,5 FPKM ; Expression significativement associée à un tissu ou à un groupe si la quantité d'ARN détectée est au moins 5 fois supérieure à la quantité détectée dans tous les autres tissus (tissu & groupe spécifiques) ou à la moyenne d'expression dans les autres tissus («surexprimés dans 1 tissu»). Seuls les gènes communs aux deux études ont été pris en compte.

A. Nombre de gènes appartenant à chaque catégorie d'expression.

B. Nombre de gènes tissu-spécifiques par organe.

2. Protéger l'information génétique et orchestrer sa duplication

Une autre fonction essentielle des épigénomes est de participer à la protection de l'intégrité du génome et de l'information génétique. Originellement, **la méthylation de l'ADN semble avoir évolué pour distinguer le génome des cellules eucaryotes des insertions parasitaires d'éléments transposables** (Slotkin & Martienssen 2007). Ces éléments parasites que sont les rétrotransposons (rétrovirus endogènes, éléments L1 ou Alu...) représentent près de 40% du génome humain et menacent son intégrité : en effet les éléments transposables actifs peuvent s'intégrer dans les séquences des gènes et entraîner des mutations importantes, ou en perturber la régulation transcriptionnelles ; et les éléments même inactifs peuvent entraîner des événements de recombinaisons entre répétitions non-alléliques, à l'origine de réarrangements chromosomiques majeurs ou de translocations (Montagna et al. 1999). L'expression des gènes encodés par les rétrotransposons (comme la reverse transcriptase) est essentielle à leur mobilité, c'est pourquoi tous les mécanismes épigénétiques assurant leur répression sont essentiels pour limiter leur dispersion.

Chez les mammifères, il est possible que les autres fonctions de la méthylation de l'ADN (et des modifications épigénétiques) aient ensuite été acquises par divergence et diversification évolutive (Colot & Rossignol 1999). **La méthylation de l'ADN joue en effet un rôle important dans la stabilité des microsatellites**, des régions répétées très sujettes à extension ou contraction lorsqu'elles sont répliquées ou transcrites. En assurant la répression transcriptionnelle de ces séquences, la méthylation de l'ADN prévient la formation de structures mutagènes dans ces régions sensibles. Additionnement à ce rôle dans la stabilisation des régions répétées, certaines études suggèrent **un rôle de la méthylation de l'ADN dans l'inhibition des événements de recombinaisons homologues**, qui peuvent être l'occasion de remaniements chromosomiques importants, et contribue à **la définition et à la protection des régions génomiques vulnérables comme les répétitions péricentromériques et centromériques** (Jaco et al. 2008)

La structure chromatinienne joue également un rôle important dans la protection de l'information génétique : **le repliement de l'ADN autour des protéines histones permet de limiter l'exposition de cette molécule au milieu oxydant et aux différents agents** génotoxiques. En effet, on a pu montrer que la molécule d'ADN « nue » subissait 300 fois plus de dommages face à un stress oxydatif que de l'ADN hétérochromatinisé, 100 fois plus que de l'ADN chromatinien, et tout de même 14 fois plus que de l'ADN nucléosomal décompacté, en collier de perles (Ljugman & Hanawalt 1992). La compaction chromatinienne protège également la molécule d'ADN des agents physiques, comme les radiations par exemple (Takata et al. 2013).

Additionnement à l'action quasi « mécanico-chimique » de la compaction chromatinienne, les activités catalytiques du système épigénétique protègent également l'information génétique. Une étude récente, s'intéressant à « la vie secrète des histones » (Rudolph & Luger 2020), a révélé **les fonctions enzymatiques d'oxydoréductase du tétramère H3-H4** (Attar et al. 2020), donnant à la chromatine un rôle important dans la gestion de la balance redox et des stress oxydatifs.

Le système épigénétique joue également un rôle dans la réparation des dommages à l'ADN.

Tout d'abord, il peut être impliqué dans la signalisation des sites endommagés : par exemple, l'un des premiers signaux des cassures double brin de l'ADN est la phosphorylation du variant d'histone γ -H2AX (Fillingham et al. 2006), qui entraînera une cascade d'événements permettant le recrutement des complexes de réparation. De même, MBD4 est impliquée dans la reconnaissance et la réparation des transitions C \rightarrow T au niveau des sites CpG (qui sont particulièrement sujets à cette mutation s'ils sont méthylés, due à la désamination spontanée des cytosines) : MBD4 reconnaît les mésappariements T : G et recrute les protéines de réparation des mésappariements comme MLH1 (Bellascosa et al. 1999, Millar et al. 2002). MBD4 joue un rôle protecteur contre l'apparition de mutations : de fait entre 26 et 43% des tumeurs du colon MSI (instabilité des microsatellites) chez l'humain présentent une mutation de MBD4 (Riccio et al. 1999).

Ensuite, **la réparation de l'ADN s'effectue dans un contexte chromatinien**, dont la compaction constitue un obstacle pour l'accès des protéines effectrices au site de dommages. Le modèle « **Prime, repair, restaure** » (Soria et al. 2012) suppose qu'après la signalisation du dommage, la structure chromatinienne doit nécessairement être dissolue pour permettre la réparation, avant d'être restaurée. Cette restauration peut impliquer l'intégration d'histones néosynthétisées, naïves, et la restauration des patrons de méthylation des CpG s'il y a lieu. Enfin, cette restauration peut également être l'occasion d'un ajout de modifications épigénétiques permettant de signaler l'intervention des voies de réparation de l'ADN, telle une cicatrice venant assurer la mémoire des dommages subis (pour revue : Smeenk & van Attikum 2013).

Pour finir, l'organisation tridimensionnelle de la chromatine soutient la **compartmentalisation de certaines voies de réparation au sein du noyau** : par exemple, les pores nucléaires sont les lieux d'adressage des dommages persistants, où se trouvent séquestrés des acteurs spécifiques de réparation (pour revue : Hauer & Gasser 2017). Plus généralement, et quel que soit le type cellulaire, **l'organisation de la chromatine soutient le bon déroulement de la mitose**. En effet, des études génétiques ont démontré que l'inactivation de certains modulateurs épigénétiques entraîne non seulement des défauts au niveau des épigénomes, mais peut également avoir des répercussions dramatiques sur la division cellulaire. Au moins dans les cellules cancéreuses, l'inactivation complète de DNMT1 entraîne l'hémiméthylation d'une partie importante des CpG mais aussi des défauts mitotiques sévères, entraînant soit la mort cellulaire suite à une catastrophe mitotique, soit l'arrêt de la prolifération en G1 pour tétraploïdie (Chen et al. 2007).

Cette étude souligne le rôle essentiel tenu par la méthylation de l'ADN pour le bon déroulement de la division cellulaire. Certains modulateurs épigénétiques, comme la protéine à doigts de zinc ZBTB4, jouent également un rôle dans le déclenchement des checkpoints mitotiques, et leur délétion est responsable de défauts mitotiques sévères (Roussel-Gervais et al. 2017). Il en est de même pour les modifications des histones : par exemple, l'histone méthyltransférase SUV39H est essentielle à l'identification épigénétique de l'hétérochromatine péri-centromérique, et sa délétion est à l'origine de

défauts de ségrégations des chromosomes entraînant une instabilité génomique importante (Peters et al. 2001).

Lors de la division cellulaire, il est en effet crucial de coordonner les événements chromatiniens avec les phénomènes cellulaires permettant la formation des deux cellules-filles : le programme de réplication de l'ADN est donc strictement défini dans le temps et dans l'espace. Ce programme résulte de l'intégration de nombreuses caractéristiques structurales de la chromatine : certaines dépendent directement de la séquence primaire de l'ADN, telles que la densité en gènes ou la richesse des séquences en AT ; d'autres vont varier d'un type cellulaire à l'autre, tels que le niveau de transcription des gènes, les modifications des histones, la méthylation de l'ADN et l'organisation tridimensionnelle de la chromatine (**FIGURE 21**). Ainsi, l'épigénome ne doit pas seulement être passivement reproduit au cours de la réplication : **les marques épigénétiques sont des actrices à part entière** de ce processus.

De façon générale, les régions actives et accessibles de l'euchromatine sont répliquées en premier lors de la phase S, et les régions réprimées de l'hétérochromatine sont répliquées plus tardivement, à la fin de la phase S. La différenciation cellulaire sera donc également l'occasion d'une réorganisation des régions répliquées précocement et tardivement, de façon concomitante à la réorganisation globale des épigénomes (Hiratani et al. 2010). Enfin, les mécanismes épigénétiques interviennent également pour **assurer la division asymétrique des cellules souches et des progéniteurs**. Durant ce processus, une cellule mère donne naissance à deux cellules filles qui possèdent le même matériel génétique mais prennent deux destinées différentes : l'une s'engagera dans un processus de différenciation tandis que l'autre renouvellera le pool de cellules souches. L'inégale répartition des marques épigénétiques garantit, dans ce cas précis, l'inéquivalence des informations transmises à chaque cellule fille et la spécificité de leur programme d'expression (pour revue : Escobar et al. 2021).

3. Altération des marques épigénétiques et pathologies humaines

Ces exemples illustrent combien les marques épigénétiques sont essentielles au bon déroulement des activités cellulaires impliquant la chromatine. Réciproquement, on imagine aisément que l'altération du système épigénétique puisse conduire à des défauts majeurs dans ces différentes fonctions essentielles. C'est pourquoi les dommages épigénétiques participent au développement de nombreuses pathologies humaines (pour revue : Robertson & Wolffe 2000). Ces pathologies peuvent être d'origine héréditaire (comme le syndrome de l'X fragile), liées à la mutation de modulateurs épigénétiques (comme par exemple les mutations de DNMT3B et l'ICF, ou de MeCP2 et le syndrome de Rett), ou encore se développer différemment au gré de l'influence de l'empreinte parentale (comme dans les syndromes de Beckwith-Wiedemann, Silver-Russel, Prader-Willi, Angelman...).

Au cours de ma thèse, je me suis intéressée aux cancers. Les cancers sont une maladie de rupture homéostatique, dans laquelle les réseaux complexes qui gouvernent cet équilibre sont systématiquement altérés, y compris les réseaux de régulation des gènes. De fait, les cellules cancéreuses montrent systématiquement à la fois une altération des fonctions chromatiniennes, avec plus de 50%

des tumeurs présentant des mutations des protéines chromatinienne, et une modification importante de leurs épigénomes. **L'identité d'origine des cellules cancéreuses est modifiée** et l'on détecte dans les tumeurs à la fois l'expression de gènes normalement réprimés et au contraire l'inhibition de gènes habituellement exprimés. Je me suis beaucoup intéressée à cette caractéristique des tumeurs, pour laquelle l'épigénétique joue un rôle important. Concernant les anomalies de mitose, plusieurs études ont montré une **dérégulation du rythme de la réplication** (*replication timing*) dans les cellules cancéreuses (**FIGURE 21**), selon un schéma conservé entre différents types de cancers (Du et al. 2019) : cette caractéristique pourrait être une nouvelle caractéristique générale du cancer. Il y a également une relation directe entre le stress réplicatif, la coordination de la réplication et l'instabilité génétique : les régions répliquées tardivement sont plus susceptibles d'accumuler des réarrangements chromosomiques en *cis*, et présentent une accumulation de mutations ponctuelles de type substitution de bases (pour revue : Briu et al. 2021). L'instabilité épigénétique est un des facteurs favorisant le développement de tumeurs : par exemple, des modèles murins hypomorphes pour *Dnmt1* affichent une hypométhylation globale de leur génome et développent des lymphomes très agressifs dès l'âge de 4 mois (Gaudet et al. 2003). Les défauts épigénétiques sont donc responsables, dans une certaine mesure, de la **mutabilité accrue des cellules cancéreuses**. De fait, le risque de cancers augmente avec l'âge ; or le vieillissement génère une accumulation d'erreurs épigénétiques qui déstabilisent peu à peu les épigénomes et l'intégrité chromatinienne. Enfin, la perte d'empreinte parentale à certains sites stratégiques (Holm et al. 2005) ou les défauts d'inactivation du chromosome X (Richart et al. 2021) peuvent être une des premières étapes de la transformation oncogénique. Par exemple, l'expression biallélique du gène soumis à empreinte Insulin-like Growth Factor-2 (IGF2) promeut la tumorigenèse en inhibant l'apoptose et est un facteur de risque pour le développement de cancers du colon (Cui et al. 2003). Plus généralement, Holms et al ont développé un modèle murin permettant d'adresser la question du rôle de la perte d'empreinte parentale dans la tumorigenèse (Holm et al. 2005). Elle a montré qu'elle pouvait être une étape très précoce de la transformation : les cellules ayant un défaut d'empreinte présentent une immortalisation spontanée. Le cancer est bien aussi une affaire d'épigénétique : prenons donc le temps de développer les aspects de cette maladie dans le chapitre suivant.

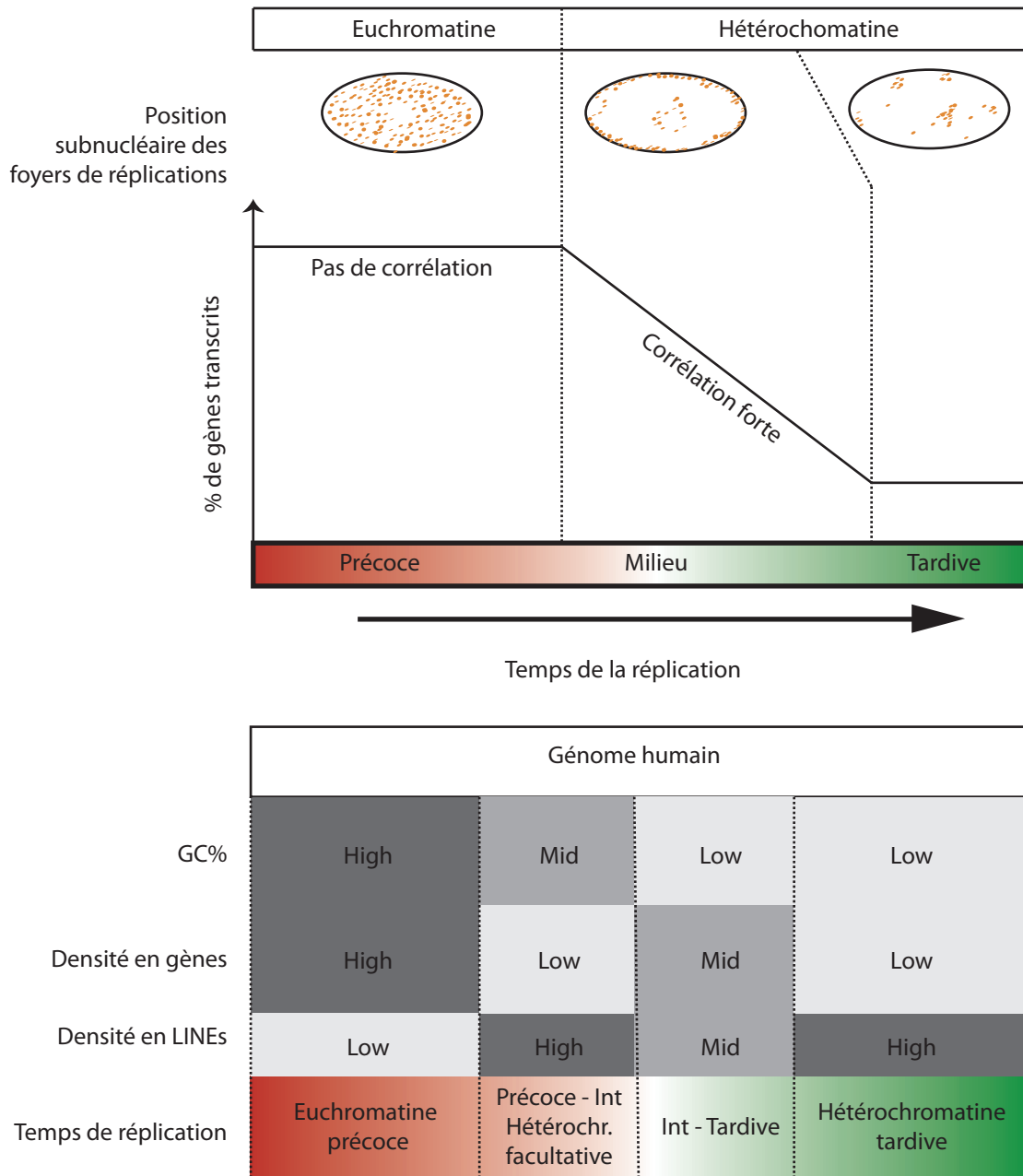


Figure 21 : Cinétique de la réplication et structure de l'information génétique

Les variations dans le timing répliatif sont corrélées avec la structure chromatinienne et le contenu de l'information génétique.

Au moment de la phase mid-S, le timing de réplication corrént avec le pourcentage de gènes exprimés. Pour les phases précices et tardives de la réplication, on n'observe pas de telle corrélation, cependant la phase précoce verra répliquer plutôt les régions d'euchromatine centrale, riches en gènes et pauvres en LINEs, riches en GC également ; tandis que la phase tardive permettra la réplication de l'hétérochromatine située en périphérie nucléaire, riche en LINEs et pauvre en GC (adapté de Rhind & Gilbert 2013)

CHAPITRE 2

Cancers et soi-modifié

Partie I

Quelles sont les caractéristiques fondamentales et permissives des cancers ?

Historiquement, le mot « tumeur » est un terme anatomique qui décrit un gonflement localisé causant une déformation d'un organe ou d'une partie du corps. Une tumeur pouvait donc être provoquée par des lésions différentes, notamment une accumulation de liquide ou une tuméfaction d'origine inflammatoire. La définition actuelle est plus restrictive et repose sur la notion **d'homéostasie** : le terme de tumeur désigne aujourd'hui une masse tissulaire ressemblant plus ou moins au tissu normal homologue, ayant tendance à persister et à croître, ce qui témoigne de son autonomie biologique.

Le cancer est un enjeu majeur en santé publique. En effet, l'incidence des cancers est en augmentation depuis la seconde moitié du XXe siècle : on peut notamment incriminer les effets conjugués de l'accroissement de l'espérance de vie et du vieillissement de la population, mais aussi l'augmentation de la prévalence de facteurs de risques (tabagisme, surpoids, sédentarité...) (Torre et al. 2012).

Les dernières décennies de recherche ont permis de faire progresser notre connaissance des mécanismes fondamentaux à l'origine des tumeurs : un petit nombre d'entre eux sont communs à tous les types de tumeurs humaines, quel que soit le tissu d'origine ou les événements initiateurs de la tumeur.. Ces caractéristiques fondamentales des cellules cancéreuses ont été théorisées en 2000 puis en 2011 par Robert Weinberg (FIGURE 22) : elles sont au nombre de huit, sont acquises progressivement au fil du développement multi-étapes des tumeurs, et déterminent collectivement l'évolution de la maladie. L'acquisition de ces 8 caractéristiques fondamentales est rendue possible par deux caractéristiques supplémentaires, dites « permissives », qui provoquent l'apparition d'un cancer ou en favorisent la progression : ce sont l'inflammation pro-tumorale et l'instabilité génétique (Hanahan & Weinberg 2000, 2011)

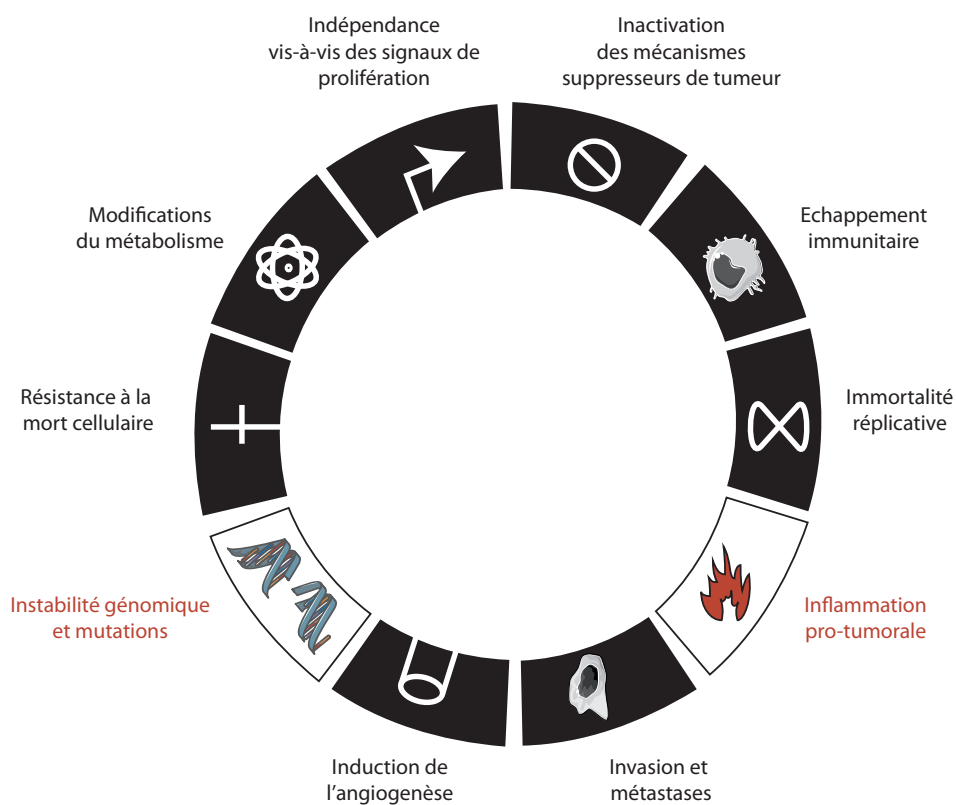


Figure 22 : Caractéristiques fondamentales et permissives des cancers

Les 8 caractéristiques fondamentales sont représentées en noir, les 2 caractéristiques permissives en rouge. (D'après Hanahan et Weinberg 2011)

1. Les cancers détournent les fonctions cellulaires pour acquérir 8 caractéristiques fondamentales

La caractéristique principale des cellules cancéreuses est avant tout cette perte d'homéostasie, c'est à dire la rupture de l'équilibre entre prolifération et mort cellulaire. Ceci est rendu possible par plusieurs mécanismes. Les tissus normaux contrôlent la prolifération des cellules grâce à l'émission parcimonieuse et régulée de signaux de prolifération (par exemple des facteurs de croissance), afin d'assurer l'homéostasie du tissu, son architecture et sa fonction. Or, contrairement aux cellules normales qui ne prolifèrent que lorsqu'elles reçoivent des signaux de prolifération extérieurs, les cellules cancéreuses acquièrent la capacité de rentrer en division cellulaire **indépendamment de tout signal de prolifération**. Pour ce faire, elles ont recours à de différentes stratégies : elles peuvent par exemple produire elles-mêmes des facteurs de croissance, pour s'auto-stimuler via un signal endocrine. Elles peuvent également utiliser les cellules du microenvironnement tumoral afin qu'elles leur fournissent ces facteurs de croissance, de façon cette fois-ci paracrine. Elles peuvent aussi élever le niveau de leurs récepteurs de facteurs de croissance ou modifier la structure de ces récepteurs pour être plus sensibles aux ligands : c'est ce qui se passe, nous le verrons, dans le cas de certains cancers du sein. **Toutes ces stratégies reposent sur l'activation de mécanismes dits « pro-oncogènes », car stimulant la croissance cellulaire.**

Les cellules cancéreuses doivent également **contourner des régulations négatives qui limitent la croissance cellulaire**, qui dépendent notamment de l'action de gènes dit « **suppresseurs de tumeurs** ». Deux exemples de gènes suppresseurs de tumeurs fréquemment inactivés dans les cancers sont les gènes *RB* et *TP53* (**FIGURE 23**) (Bikhart & Sage 2008).

La protéine RB est responsable de l'intégration de nombreux signaux, notamment extracellulaires, et gouverne la décision d'entrer ou non en cycle cellulaire : les cellules cancéreuses déficientes pour cette voie acquièrent ainsi la capacité d'entrer en cycle quel que soient les signaux extrinsèques. De façon complémentaire, la protéine TP53 est chargée de l'intégration de nombreux signaux intracellulaires : notamment, elle assure l'impossibilité à une cellule présentant un trop haut niveau de stress (tels que des dommages à l'ADN ou une carence en métabolites) d'entrer en cycle tant que sa situation ne s'est pas stabilisée. Si les dommages sont trop importants, l'activation prolongée de TP53 entraînera un programme d'apoptose permettant l'élimination des cellules défectueuses. L'inactivation de TP53 est un mécanisme fréquent d'échappement des cellules tumorales à ce contrôle : environ 60% des cancers ont une mutation de leur gène P53.

Par ailleurs, et outre les régulations négatives induites par les suppresseurs de tumeurs, une autre limitation importante de la prolifération cellulaire est assurée dans les cellules normales par les adhésions de contact inter-cellules, qui limitent la prolifération. **Les inhibitions de contact sont abolies dans la plupart des tumeurs**, permettant aux cellules cancéreuses de continuer à proliférer bien au-delà de l'homéostasie du nombre de cellules.

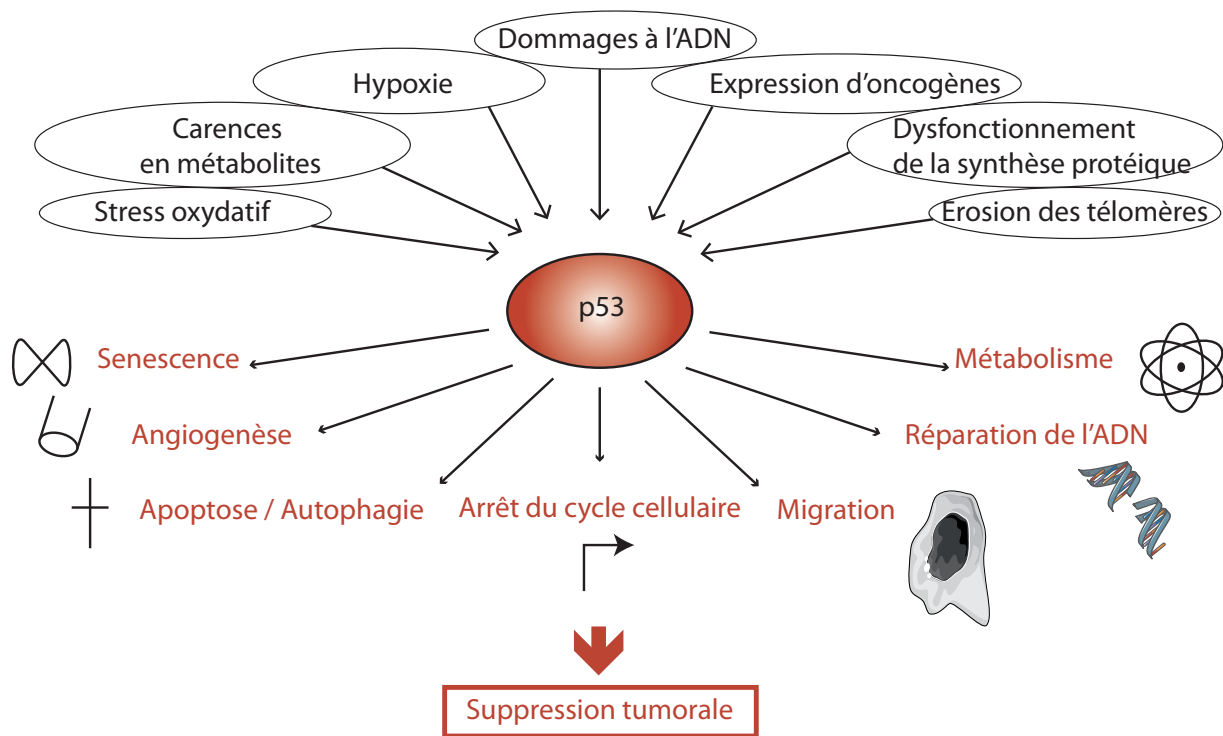


Figure 23 : Inactivation des gènes suppresseurs de tumeurs : TP53, un master régulateur

TP53 est une protéine «hub», responsable de l'intégration de nombreux stress cellulaires et en particulier ceux touchant les défauts chromatiniques. Véritable «gardien du génome», il joue un rôle de suppresseur de tumeur en engageant les cellules défectueuses vers les voies de réparation et en limitant leur développement si les réparations échouent. Les cellules tumorales échappent à la surveillance de TP53 en l'inactivant.

De plus, **les cellules cancéreuses acquièrent un potentiel de réplication illimité**, assurant leur «immortalité» virtuelle. Dans les cellules saines, la limitation du potentiel réplicatif est assurée par l'érosion progressive des télomères, à chaque division. Cette érosion définit une « horloge moléculaire », limitant le nombre de cycles réplicatifs que peut connaître une cellule : en effet, lorsque les télomères sont trop courts, ils ne peuvent plus remplir leur fonction de protection des extrémités chromosomiques ce qui déclenche une crise cellulaire, au cours de laquelle la cellule meurt. Pour contrecarrer ce mécanisme, plus de 90% des cellules immortelles réactivent le gène codant la télomérase, une enzyme capable d'assurer la maintenance de la longueur des télomères au cours des cycles de réplication.

De l'autre côté de la balance homéostatique, **les cellules cancéreuses sont aussi capables d'échapper à la mort par apoptose**, ce mécanisme qui assure dans les tissus sains un obstacle endogène à l'accumulation de cellules anormales en permettant leur élimination. Les cellules cancéreuses déploient différentes stratégies pour échapper à l'apoptose : le plus commun est l'inactivation de gènes suppresseurs de tumeurs tels que TP53, comme nous l'avons vu. Alternativement, certaines cellules cancéreuses peuvent jouer sur la balance entre signaux pro- et anti-apoptotique, pour éviter l'activation de ces voies.

Perte d'homéostasie et immortalisation sont les deux caractéristiques de la première étape de la cancérogenèse : **l'initiation, ou transformation**. Cette étape peut ensuite être suivie d'une phase de

promotion cancérogène, qui pérennise la rupture homéostatique et amplifie les anomalies au cours des divisions cellulaires. Le nombre de cellules cancéreuses augmente au cours de la promotion, d'abord de façon lente puis exponentielle : cette progression fait passer la maladie du stade infraclinique (absence de symptômes) au stade clinique. Dans le cas des cancers solides, le développement de la masse de cellules cancéreuses donnent naissance à une tumeur, en général à partir de 10^9 cellules. A partir de cette étape, on parlera de **tumorigenèse** : la tumeur grossit et connaîtra une évolution locale, régionale voire métastatique. Cette étape s'accompagne de la nécessaire coopération du stroma tumoral, et de l'acquisition de capacités supplémentaires par les cellules cancéreuses.

Pour assurer sa croissance et son approvisionnement en oxygène et en nutriments, **la tumeur acquiert la capacité d'induire et de stimuler l'angiogenèse**, c'est-à-dire le développement de nouveaux vaisseaux sanguins venant irriguer la masse tumorale. Pour ce faire, on observe notamment la dérégulation de la balance entre signaux pro- et anti-angiogéniques : de nombreuses tumeurs surexpriment les facteurs pro-angiogéniques VEGF (*vascular endothelial growth factor*) et FGFs (*fibroblast growth factors*), et diminuent en parallèle l'expression de facteurs inhibiteurs.

De fait, la croissance incontrôlée des cellules tumorales oblige celles-ci à opérer de nombreux **ajustements du métabolisme énergétique**, afin de soutenir la croissance cellulaire. L'une des adaptations métaboliques des cellules cancéreuses est l'inhibition de la phosphorylation oxydative, favorisant la production d'ATP via la glycolyse (Warburg et al. 1956). Ce mécanisme est appelé « effet Warburg », et permet d'augmenter la production d'intermédiaires métaboliques nécessaires à la biosynthèse de nombreuses molécules (Vander Heiden et al. 2009), compensant ainsi la moindre efficacité de production d'ATP.

Un moment clé de l'évolution de la maladie tumorale est **l'acquisition des capacité d'invasion et de métastase** par certaines cellules cancéreuses, dites pionnières. Ce passage marque une étape cruciale, car **les métastases sont responsables de 90% des morts par cancers** (Sporn et al. 1996). Des mécanismes complexes soutiennent ces processus, et déterminent notamment la transition entre des états cellulaires plus épithéliaux et plus mésenchymateux (FIGURE 24), au cours desquelles les cellules cancéreuses prennent différents états hybrides entre un phénotype très solidaire de son tissu et un phénotype plus autonome et capable de se déplacer (Blanpain et al. 2018). De nombreuses recherches soulignent aujourd'hui le rôle crucial joué par le microenvironnement tumoral dans l'acquisition de la capacité tumorale à envahir le tissu avoisinant et à métastaser. On retiendra notamment le dialogue déterminant entre macrophages associés à la tumeur (producteurs d'EGF, qui stimule l'intravasation des cellules tumorales) et cellules cancéreuses (productrices de CSF-1, qui stimule les macrophages) dans le cas des cancers du sein métastatiques (Qian & Pollard 2010).

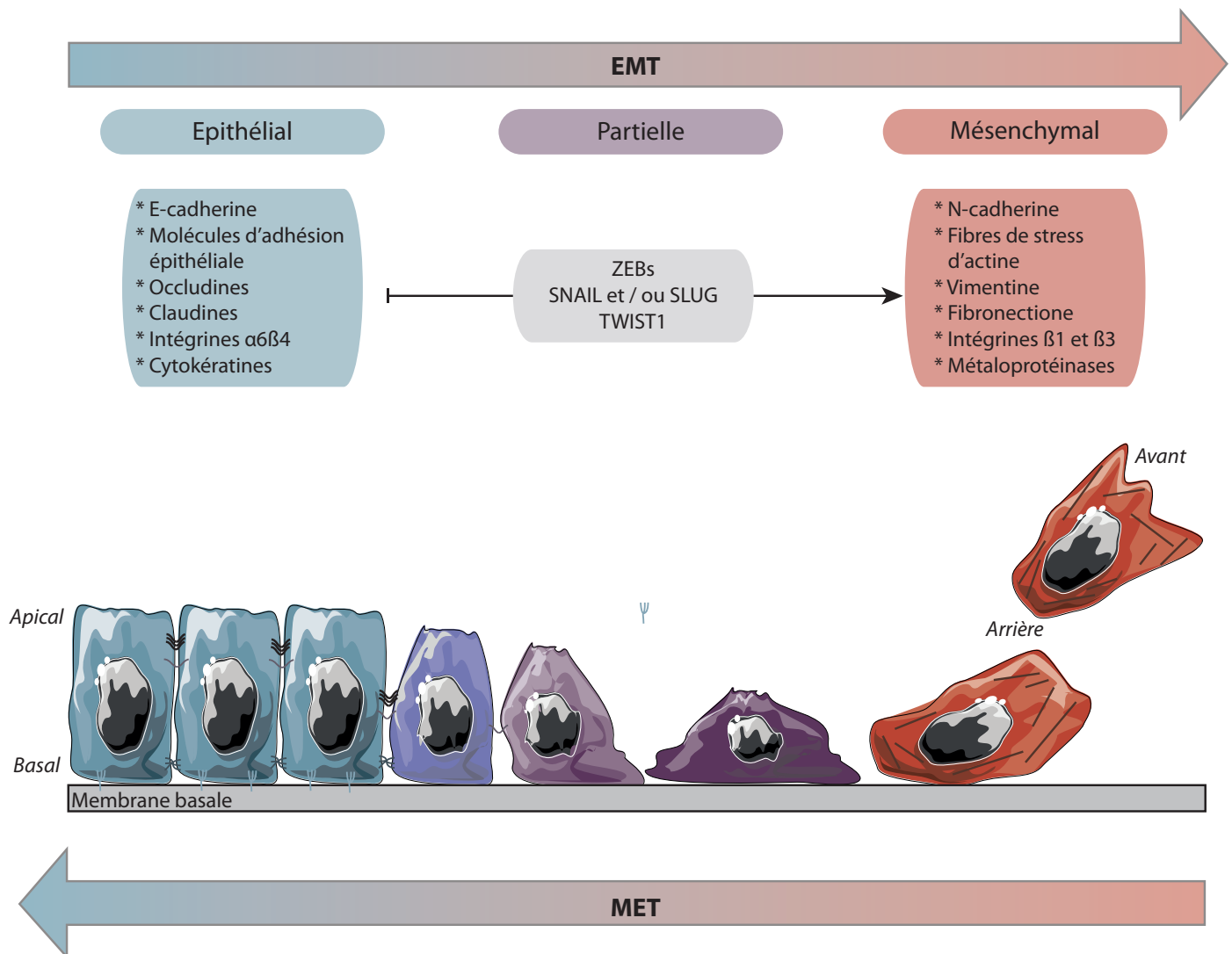


Figure 24 : Etats de transition de l'EMT et progression métastatique des cellules carcinomateuses

Lors de la dissémination métastatique, les cellules cancéreuses épithéliales (à gauche) peuvent évoluer vers un phénotype d'EMT partiel (au milieu) ou un phénotype totalement mésenchymateux. Durant l'EMT, les cellules épithéliales perdent leur polarité apico-basale et leurs adhésions inter-cellulaires ainsi qu'avec la membrane basale. L'EMT est régulée par différents facteurs, dont certains facteurs de transcription (ZEB1, SNAIL, TWIST1...) et des modifications épigénétiques permettant l'expression de molécules spécifiques. Le processus inverse à l'EMT s'appelle MET (flèche du bas). (Adapté de Dongre et al. 2019)

Enfin, **la dernière caractéristique fondamentale des tumeurs est la capacité** à échapper au système immunitaire. En effet, la surveillance exercée par les cellules immunitaires est théoriquement suffisante pour identifier et éliminer les cellules anormales, voire cancéreuses : dans cette logique, il apparaît nécessaire pour les tumeurs de parvenir à passer sous les radars du système immunitaire et de limiter leur élimination. Pour ce faire, les cellules tumorales peuvent anergiser (inactiver) les lymphocytes T infiltrant (TILs) et les cellules Natural Killers (NKs), en émettant des signaux immunosuppresseurs (Yang et al. 2010, Shields et al. 2010). Des mécanismes indirects peuvent aussi rentrer en jeu, comme le recrutement de cellules immunitaires de type lymphocytes T régulateurs (Treg) ou *myeloid-derived suppressor cells* (MDSCs) par la tumeur, qui exerceront une action immunosuppressive sur les autres cellules immunitaires (Mougiakakos et al. 2010, Ostrand-Rosenberg & Sinha 2009). Dernièrement, cet aspect de la biologie des tumeurs a été propulsé sur le devant de la scène grâce aux développements de stratégies thérapeutiques exploitant justement l'immunité anti-tumorale.

2. Deux caractéristiques permissives favorisent le développement oncogénique

Ces huit caractéristiques sont acquises au cours de l'oncogenèse grâce à deux caractéristiques permissives essentielles.

1. Le génome des cellules cancéreuses est caractérisé par une mutabilité supérieure aux cellules saines

Une caractéristique permissive essentielle est l'acquisition d'un certain degré d'instabilité génomique dans les cellules cancéreuses, capable de générer de multiples altérations aléatoirement à travers le génome. Ces mutations sont un des mécanismes permettant l'activation de proto-oncogènes ou l'inhibition de gènes suppresseurs de tumeurs ; on notera cependant que des mécanismes épigénétiques peuvent également permettre la répression ou l'activation stable de ces gènes.

La mutabilité des cellules cancéreuses est rendue possible par **l'acquisition de défauts dans les composants de la machinerie de maintenance du génome**, impliquant : le système de senseurs capables de détecter les dommages à l'ADN, d'activer les voies de réparations et de bloquer la prolifération cellulaire tant que les lésions sont trop importantes (encodés par les gènes «portiers» ou **gate keeper genes** : APC, RB1, TP53...) ; les enzymes directement responsables de la réparation de l'ADN (encodés par les gènes «soignant» ou **care taker genes** : BRCA1/2, PARP1, MLH1...) ; et enfin les systèmes cellulaires impliqués dans la gestion des molécules génotoxiques. Le type et la nature des altérations génomiques varient beaucoup selon les types de tumeurs, cependant on retrouve dans la vaste majorité des cancers un certain degré d'instabilité génomique. En effet, les défauts de maintenance et de réparation du génome confèrent aux cellules cancéreuses un avantage sélectif, en permettant à la tumeur d'accélérer le rythme d'acquisition des mutations. Ainsi, cette instabilité génomique est l'un des mécanismes permissif essentiel permettant aux cellules tumorales de détourner les fonctions cellulaires normales et d'acquérir les caractéristiques fondamentales que nous avons décrites.

2. L'inflammation du tissu tumoral soutient la croissance déséquilibrée des cellules cancéreuses

L'examen histologique d'un tissu tumoral indique clairement que certaines tumeurs sont densément infiltrées par une variété de cellules immunitaires, et montrent toutes les marques classiques d'une réponse inflammatoire. La réponse inflammatoire est déclenchée à chaque fois qu'un tissu subit une agression, ce qui est le cas lors du développement des lésions tumorales. Pendant longtemps, cette réponse inflammatoire a été interprété comme le signe de la tentative du système immunitaire d'éradiquer la tumeur, et de fait, de nombreux éléments de preuves soutiennent le fait **qu'il y a bien dans de nombreuses tumeurs une réaction immune anti-tumorale au site inflammé**, parallèlement aux efforts déployés par les cellules cancéreuses pour échapper à la surveillance immunitaire.

Cependant, il est aujourd'hui clair **que le statut inflammatoire du tissu tumoral a paradoxalement un effet pro-tumoral**, et peut également soutenir la progression de la maladie. En effet, la réaction inflammatoire entraîne la production de nombreuses molécules qui se trouvent alors disponibles pour

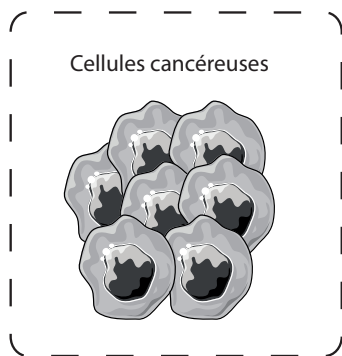
la tumeur, parmi lesquelles des facteurs de croissance et de survie, des facteurs pro-angiogéniques ou de remodelage de la matrice extracellulaire. L'inflammation favorise ainsi l'acquisition de certaines caractéristiques importantes des cellules cancéreuses, telles que la prolifération, l'échappement à la mort cellulaire, l'invasion et l'angiogenèse. Par exemple, dans les cancers du sein métastatiques, le dialogue réciproque entre macrophages associés à la tumeur (producteurs d'EGF, qui stimule l'invasion des cellules tumorales), et cellules cancéreuses (productrices de CSF-1, qui stimule les macrophages) est déterminant pour l'évolution de la tumeur (Qian & Pollard 2010). De plus, les cellules immunitaires activées présentes au sein du tissu tumoral peuvent également être responsables de la production de composés générateurs de stress oxydatif et donc favorisant la mutabilité du génome des cellules cancéreuses. Ainsi, l'inflammation peut être vue comme un facteur facilitant l'acquisition des caractéristiques fondamentales des cellules tumorales, et le système immunitaire se trouve donc paradoxalement impliqué à la fois dans le combat contre les cellules tumorales et dans leur soutien.

3. Les tumeurs sont des systèmes hétérogènes

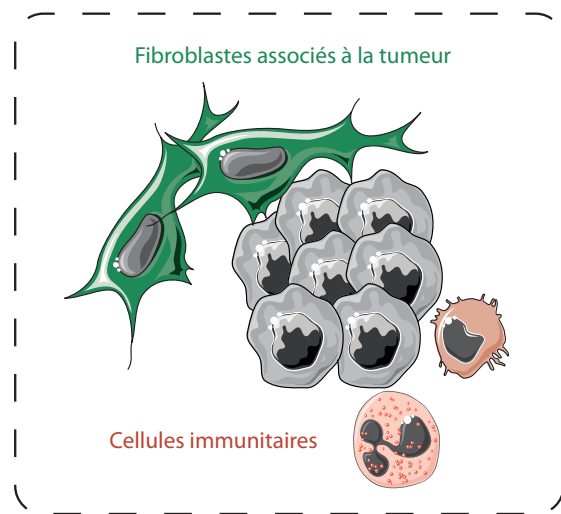
Au cours des dernières décennies, en effet, notre appréhension des cancers s'est complexifiée, en envisageant non plus les tumeurs comme une simple masse de cellules cancéreuses mais comme de véritables systèmes organiques, s'appuyant sur le fonctionnement coordonné de multiples types cellulaires qui communiquent entre eux et s'influencent mutuellement (FIGURE 25). Cette complexité du tissu tumoral est conceptualisée par le terme de « **microenvironnement tumoral** », qui désigne l'ensemble des cellules somatiques non transformées mais faisant partie intégrante de la masse tumorale, ensemble dynamique construit tout au long des multiples étapes de l'oncogenèse.

De ce fait, l'analyse dite en « bulk » des caractéristiques génétiques, transcriptomiques et épigénétiques d'une tumeur reflètera un profil moyen, auquel contribuent des types cellulaires variés. Certaines approches bioinformatiques ont cherché à tirer profit de ce biais, en identifiant **des signatures transcriptomiques dont l'expression traduirait la présence de ces cellules somatiques associées à la tumeur**, de façon semi-quantitative (Bindea et al. 2013 ; Brecht et al. 2016). Ces approches reposent sur l'identification de gènes spécifiques des types cellulaires concernées (fibroblastes, neutrophiles, lymphocytes T régulateurs...), présentant les propriétés que nous avons décrites précédemment. Cependant ces études se sont appuyées sur l'identification de signature à partir des homologues « sains » des cellules somatiques, c'est-à-dire ne participant pas à ce réseau d'interactions réciproques avec les cellules tumorales, qui inéluctablement modifient leur identité et leurs signatures, ce qui constitue un artefact importante. Pour diminuer l'influence de ce paramètre, il est possible de dissocier les cellules composant la tumeur et de **trier les différentes populations** sur la base de marqueurs. Cependant, ces techniques reposent sur l'identification *a priori* de marqueurs spécifiques, et la régression infinie vers l'identification de sous-populations toujours plus spécifiques est aussi vertigineuse que coûteuse. De nos jours, les **analyses de séquençage à haut débit dites en cellule unique** (*single-cell*) nous autorisent à décortiquer plus en détail le microenvironnement tumoral, et à arbitrer entre les modifications imputables aux cellules cancéreuses et celles impliquant les cellules somatiques associées à la tumeur.

Tumeur = cellules cancéreuses clonales



Tumeur = cellules cancéreuses clonales + microenvironnement



Tumeur = cellules cancéreuses hétérogènes (génétique & différenciation) + microenvironnement

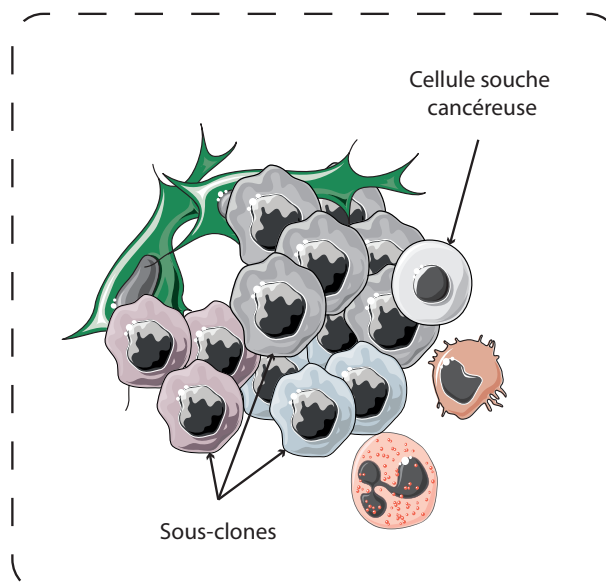


Figure 25 : Evolution conceptuelle de la notion de tumeur

Les modèles tumoraux ont gagné en complexité, en intégrant progressivement l'action pro- ou anti-tumorale du microenvironnement, et l'hétérogénéité génétique et épigénétique des cellules tumorales

L'étude des interactions des cellules cancéreuses avec leur environnement pourrait être une des clés permettant d'expliquer la cancérogenèse. Une théorie alternative à la vision déterministe de l'établissement des identités cellulaires serait de considérer que les gènes s'expriment de façon probabiliste et **que l'apparition des caractéristiques de différenciation est un phénomène aléatoire**. Dans ce cadre théorique, les patrons d'expression des gènes ne sont stabilisés qu'a posteriori, lors d'un processus de sélection par l'environnement cellulaire des cellules qui expriment les phénotypes adéquats : **seules les cellules capables d'établir des interactions et des communications stables et stabilisatrice avec cet environnement seraient conservées**. Ainsi, la cancérisation pourrait être liée à la perte ou la modification des interactions cellulaires, conduisant à la rupture de cet équilibre tissulaire et à la mise en place d'un nouvel équilibre entre cellules cancéreuses et cellules du microenvironnement tumoral, toutes présentant des modifications du point de vue de l'expression des gènes afin d'assurer ce nouvel équilibre.

Le tissu tumoral est donc un tissu complexe, intégrant de façon organique différents types cellulaires dont les cellules cancéreuses ne sont qu'une des composantes. Mais pour rendre ce tableau plus juste, il faut prendre en compte également **l'hétérogénéité intrinsèque des cellules cancéreuses**. Dès le 19^e siècle, Rudolph Virchow, l'un des pères de la physiopathologie moderne avait relevé l'existence de différents phénotypes cellulaires au sein d'une même tumeur (Brown & Fee 2006). Grâce au développement des techniques d'analyse à haut débit, cette hétérogénéité intratumorale a été extensivement caractérisée dans la seconde moitié du 20^e siècle, confirmant l'existence de sous-populations distinctes de cellules cancéreuses.

En s'appuyant sur des analyses mutationnelles de tumeurs, on peut identifier différentes sous-populations de cellules cancéreuses (FIGURE 26). Ces populations sont apparentées, car elles partagent certaines altérations génétiques ; mais elles présentent également des sets de mutations génétiques spécifiques, témoignant de leur évolutivité. Ainsi, en accélérant le rythme d'acquisition des mutations, l'oncogenèse favorise l'apparition et la sélection de sous-clones tumoraux génétiquement distincts au cours de l'évolution de la tumeur (FIGURE 26) (Kandoth et al. 2015, Vignot et al. 2013, Ding et al. 2010, Hinohara & Polyak 2019). **Ces clones peuvent s'influencer mutuellement** à travers des interactions paracrines ou de contact (Zhang et al. 2015) ; ils peuvent également présenter des **vulnérabilités ou des résistances spécifiques** aux traitements anti-cancéreux ; et peuvent enfin participer aux **processus métastatiques** en favorisant l'apparition de clones plus mésenchymateux. Les pressions de sélections qui s'exercent sur la tumeur, et auxquelles contribuent des facteurs intrinsèques tels que le système immunitaire, mais aussi les traitements anti-cancéreux qui seront administrés, favorisent le développement de certains sous-clones et influencent la composition de la tumeur, qui évoluera au fil de l'histoire de la maladie. L'hétérogénéité clonale de la tumeur primaire et de ses métastases est donc un paramètre essentiel pour comprendre l'écosystème tumoral, qui peut avoir d'importantes conséquences sur la prise en charge de ces cancers et sur l'utilisation de combinaisons thérapeutiques appropriées.

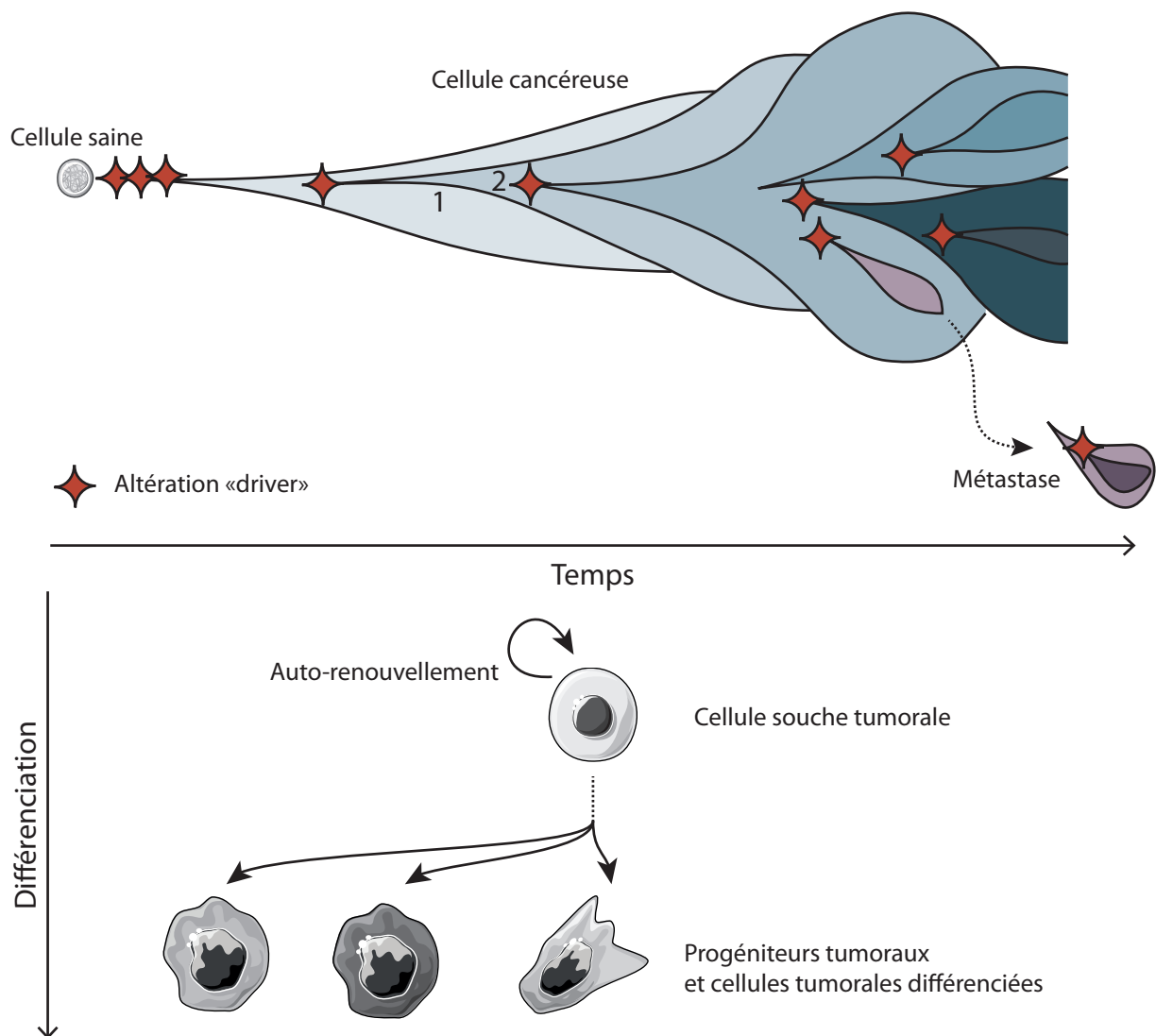


Figure 26 : Evolution clonale des tumeurs

Inspiré de (Polyak et al. 2019)

Le deuxième type de modèle décrivant l'évolution ramifiée des tumeurs au cours de l'oncogenèse s'appuie sur la découverte de **cellules souches cancéreuses** (CSCs), une sous-population de cellules cancéreuses aux propriétés mimant celles des cellules souches adultes. Ce modèle a été initialement démontré dans le cas des leucémies humaines (Bonnet & Dick 1997), avant d'être généralisé à de nombreuses tumeurs solides, y compris les cancers du sein (Al-Hajj et al. 2003) ; et postule qu'à l'instar des tissus différenciés, les tumeurs ont elles-aussi une organisation hiérarchique. Des expériences de **xénogreffes dérivées de tumeurs** de patientes ont défini de façon fonctionnelles les CSCs : ce sont des cellules cancéreuses possédant la capacité de générer une nouvelle tumeur après transplantation dans un modèle de souris immunodéficente (appelées en conséquence **tumor-initiating cells**, Cho & Clarke 2008, Ginestier et al. 2007). La nouvelle tumeur présentera généralement les mêmes caractéristiques que la tumeur d'origine, récapitulant l'hétérogénéité phénotypique des cellules cancéreuses de la tumeur primaire. Les CSCs peuvent être isolées sur la base de différents marqueurs (par exemple, CD44, CD133 ou ALDH), cependant leur identification précise reste difficile du fait de la variabilité de ces marqueurs selon les types tumoraux. Ces expériences ont permis d'ajouter une nouvelle dimension à l'hétérogénéité tumorale, composée de cellules cancéreuses différenciées mais aussi de CSCs en proportion variable.

Les CSCs présentent un degré de différenciation faible, possèdent des propriétés d'auto-renouvellement et sont capables de générer les cellules plus différenciées formant le corps de la tumeur. De nombreuses recherches ont souligné **le lien essentiel entre état souche et transition épithélio-mésenchymateuses (EMT)** : en effet, l'induction de l'EMT dans certains modèles entraîne l'apparition de cellules possédant de nombreuses caractéristiques de cellules souches, y compris la capacité d'auto-renouvellement (Main et al. 2008, Morel et al. 2008). De plus, les CSCs ont une plasticité accrue, pouvant dans certains contextes générer différents progéniteurs cancéreux mésenchymateux ou épithéliaux, eux-mêmes capables de générer leurs descendances respectives (FIGURE 26). Ainsi, la présence de CSCs capables de donner naissance à différents lignages de cellules cancéreuses influence la composition finale de la tumeur. Enfin, il est désormais établi que les cellules cancéreuses aux propriétés de CSCs sont particulièrement **résistantes aux chimiothérapies anti-cancéreuses standards**, et seraient donc responsables de la récurrence de certaines tumeurs (Ginestier et al. 2017). Les CSCs pourraient présenter un état de dormance et persisteraient sous une forme latente avant de récidiver des années (voire des décennies) après les premiers traitements, représentant de fait une épée de Damoclès suspendue à vie au-dessus des patients atteints de cancer. Par conséquent, le développement de thérapies ciblées sur les CSCs, notamment en visant les voies de signalisation spécifiques de ces cellules ou en forçant leur différenciation, suscite un grand intérêt.

Evidemment, **ces deux modèles de l'évolution des tumeurs (par évolution clonale aléatoire et différenciation de CSCs) ne sont pas exclusifs**. Les CSCs sont également susceptibles de subir des mutations génétiques spécifiques, générant ainsi différentes populations cellulaires composées de

différents sous-clones de CSCs et de leurs descendance respectives. De plus, il faut également compter avec la plasticité des cellules cancéreuses : certaines études suggèrent que les cellules tumorales les plus différenciées pourraient acquérir des propriétés de cellules souches à travers des processus de dédifférenciation.

Quoi qu'il en soit, l'existence de cette hétérogénéité intra-tumorale (qu'elle soit d'origine génétique, cellulaire ou épigénétique) pose un problème majeur pour la décision thérapeutique, rendant difficile la validation de biomarqueurs oncologiques et la prédiction de résistances thérapeutiques. L'accumulation de données nombreuses et multidimensionnelles a donc représenté une étape importante dans la compréhension de ces maladies : certaines de ces bases de données ont été très utiles à mon projet.

4. Les consortia de génétique et épigénétique des cancers

Plusieurs projets se sont donnés pour objectif de caractériser par des méthodes de séquençage à haut débit les tissus normaux et les tumeurs. Ces consortia regroupent de multiples centres, parfois internationaux, et mettent à disposition de la communauté scientifique les données collectées. Dans le cadre de ma thèse, j'ai exploité ces ressources d'une grande richesse pour caractériser les profils d'expression et les paysages épigénétiques des cellules normales et cancéreuses.

Du côté des cellules et tissus sains, j'ai utilisé des données extraites du projet **Encyclopedia of DNA Element (ENCODE)**. Ce projet de recherche public, lancé en 2003 par le *National Human Genome Research Institute*, vise à identifier tous les éléments fonctionnels du génome humain : dans ce but, le consortium a généré un grand nombre de données sur 147 types cellulaires et tissus différents, incluant des informations relatives aux régions transcrites, à la structure chromatinienne et aux modifications des histones, aux sites de méthylation de l'ADN, ainsi qu'aux sites de liaisons à des facteurs de transcription.

J'ai également exploité les données du projet **Genotype-Tissue Expression (GTEx)**, qui propose des données de RNA-seq générées à partir de 54 tissus différents collectés sur plus de 1000 individus, dont l'analyse permet de définir les signatures transcriptionnelle spécifique à chaque tissu humain. Ces données m'ont permis d'établir le profil d'expression des tissus sains, pour les comparer à ceux obtenus à partir des échantillons tumoraux.

Concernant les tumeurs, j'ai utilisé principalement deux grands projets, l'un analysant des tissus tumoraux extraits de cancers humains, le second des lignées cellulaires de cellules cancéreuses humaines.

Ce premier projet est le **Tumor Cancer Genome Atlas (TCGA)**, un consortium multi-institutionnel américain dont l'objectif est l'analyse moléculaire extensive de 10 000 tumeurs primaires couvrant 30 types de cancer. Ce projet à grande échelle permet de répertorier les mutations oncogéniques, les mécanismes épigénétiques influençant la tumorigenèse, de définir des sous-types clinique pertinent pour l'établissement d'un pronostic et la décision thérapeutique, et de développer de nouvelles thérapies anti-cancers. L'une des composantes importantes du TCGA est le développement d'une infrastructure

fournissant l'accès aux données génomiques via le TCGA Data Portal, permettant aux chercheurs et chercheuses du monde entier de réaliser et de valider d'importantes découvertes. Plus d'une centaine d'articles ont été publiés par des auteurs ne participant pas au consortium mais travaillant directement à partir des données du TCGA.

Le second projet est le **Cancer Cell Lines Encyclopedia (CCLE)**, une collaboration lancée en 2008 entre le Broad Institute et Novartis et intégrée depuis 2018 au projet DepMap (Broad Cancer Dependency Map Project). L'objectif de ce projet est d'intégrer des données moléculaires extensives sur un large panel de modèles cellulaires de cancers humains avec leurs profils de réponses pharmacologiques, afin d'identifier sur ces lignées de cellules des vulnérabilités thérapeutiques qui pourraient être transposées à la clinique. Ce projet intègre plus de 1000 lignées cellulaires de divers cancers humains, dont par exemple 76 lignées de carcinomes du sein, pour lesquels sont disponibles des données d'expression des gènes, d'altérations du génome, des données épigénétiques (microARN, méthylation de l'ADN et modifications des histones H3), et des données de protéomiques, le tout mis en relation avec le profil de sensibilité à plusieurs centaines de drogues différentes (projet GDSC2, *Genomic of Drug Sensitivity in Cancer*), mais aussi des données de survie après un crible perte de fonction (par siRNA ou CRISPR/Cas9, projet Achilles) contre plus de 500 cibles, permettant d'identifier des dépendances et potentiellement de nouvelles cibles thérapeutiques (Gandhi et al. 2019, Meyers et al. 2017, Corsello et al. 2017).

Partie II

Quel est le rôle de l'épigénétique dans la progression oncogénique ?

1. Les altérations épigénétiques participent à la transformation

Bien souvent, la simple analyse de l'architecture nucléaire d'une cellule cancéreuse suffit au pathologiste pour poser son diagnostic. En effet, **les noyaux des cellules cancéreuses présentent une configuration chromatinienne anormale**, que l'on peut deviner à la taille du nucléus, aux limites irrégulières du noyau ou encore à la présence de régions granuleuses et dispersées, sombres en hématoxyline, de chromatine très dense (**FIGURE 27**). Ces caractéristiques structurales du noyau des cellules cancéreuses traduisent les profondes altérations épigénétiques qui ont lieu tout au long de l'oncogenèse et de la progression de la maladie. Très précocement au cours de la transformation, on note dans les cellules cancéreuses un affaiblissement de la définition des domaines chromatinien : les barrières délimitant les domaines d'euchromatine et d'hétérochromatine se font plus floues (Baylin & Jones 2011). Il en résulte une altération du contrôle de fonctions chromatinien essentielles, comme la réplication de l'ADN ou le contrôle de l'expression des gènes (pour revue Michalak et al. 2019). Dans cette section, nous présenterons les différentes anomalies épigénétiques qui accompagnent la transformation et le développement des cellules cancéreuses.

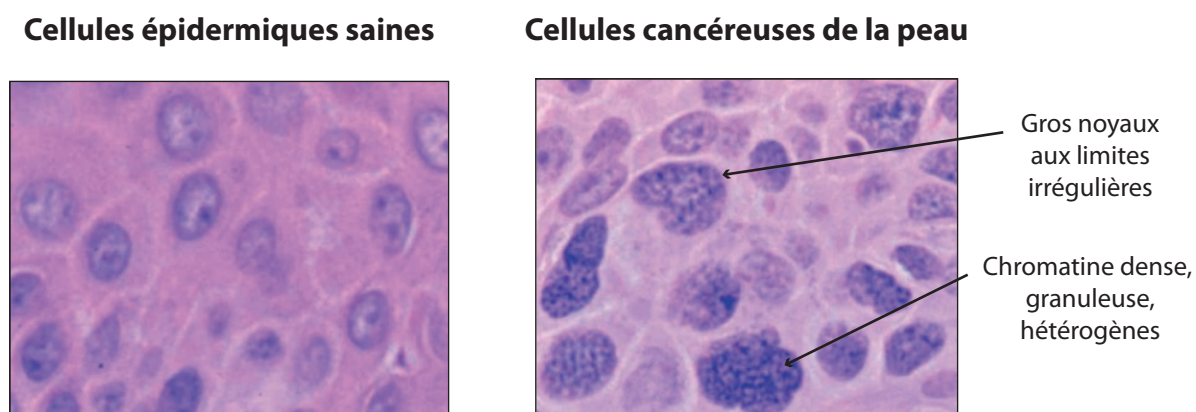


Figure 27 : Changements de la structure chromatinienne des cellules cancéreuses

Ces deux images proviennent d'un patient atteint d'un carcinome de la peau. Le panneau de gauche montre des cellules épidermiques saines, celui de droite des cellules cancéreuses de ce même patient, au même grossissement microscopique. La chromatine apparaît en violet, dû à son affinité pour l'hématoxyline. (Adapté de Baylin & Jones 2016)

Dès le début des années 70, **on relève une caractéristique moléculaire commune à la majorité des tumeurs** : celles-ci présentent **un déficit en 5mC** significatif (20-60%) par rapport aux tissus sains (Lapeyre & Becker 1979). Cette perte de méthylation globale se retrouve dans différentes régions génomiques incluant les éléments répétés, les rétrotransposons, les promoteurs pauvres en CpG, les introns et les déserts de gènes (**FIGURE 28**). **L'hypométhylation du génome contribue à l'instabilité génomique** par différents mécanismes. Tout d'abord, elle favorise les phénomènes de recombinaison mitotique, à l'origine de délétions et translocations ainsi que de réarrangements chromosomiques plus larges. Par exemple, de nombreux cancers humains (notamment les tumeurs du sein et de l'ovaire) présentent à la fois de nombreuses translocations inégales, à l'origine d'une perte d'hétérozygotie, et une hypométhylation sévère des séquences satellites péri-centromérique (Yeh et al. 2002, Esteller et al. 2008, Eden et al. 2003). En effet, la déméthylation des régions satellites favoriserait leur lésion et

recombinaison, et la déméthylation des régions centromériques peut provoquer des aneuploïdies (Karpf et al. 2005, Xu et al. 1999). Enfin, l'hypométhylation des transposons est à l'origine de leur réactivation et de leur translocation à d'autres loci, ce qui fragilisera davantage le génome (Howard et al. 2008).

Plus localement, de nombreux gènes vont se trouver différemment méthylés au cours de la transformation cancéreuse (FIGURE 28). L'hyperméthylation locale des promoteurs des gènes est également considérée comme une caractéristique générale des cancers qui apparaît relativement tôt au cours de la tumorigenèse, et est même fréquemment déjà présente dans les cellules pré-malignes (Esteller et al. 2000, Lecchesi et al. 2008).

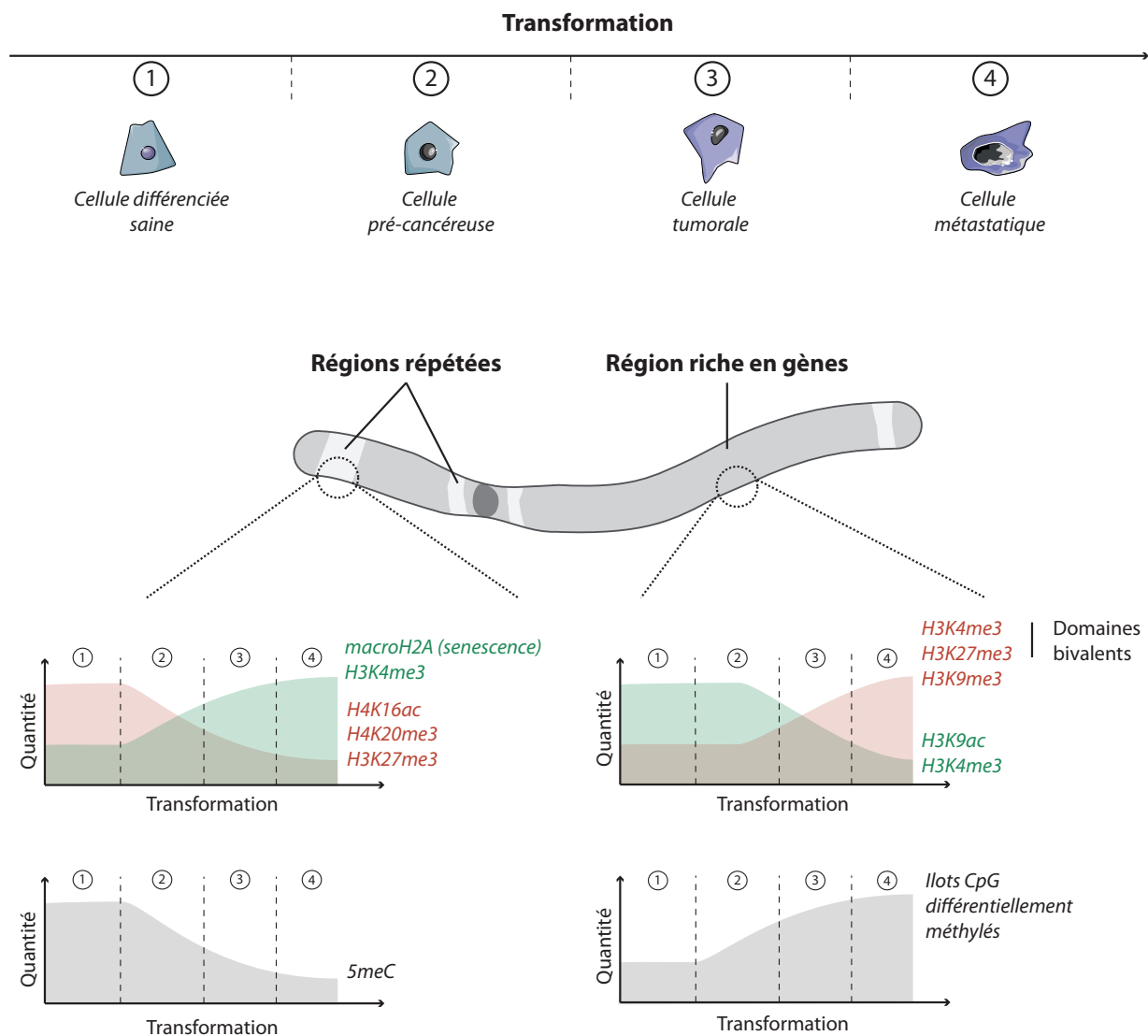


Figure 28 : Cinétique des anomalies épigénétiques au cours de la progression tumorale

Les anomalies épigénétiques sont une caractéristique systématique des tumeurs, et leur évolution accompagne la progression de la maladie. Les principales altérations des marques chromatinienne et leurs variations globales aux différentes étapes de l'évolution des tumeurs ont été représentées.

Elle affecte des gènes impliqués dans tous les processus cellulaires (Esteller et al. 2005), en particulier des gènes suppresseurs de tumeurs. Différentes approches ont cherché à quantifier le nombre de gènes touchés par cette altération : on compte entre 100 et 400 promoteurs présentant un îlot CpG hyperméthylés par tumeur, ce résultat pouvant varier selon le type tumoral. Ce chiffre reste néanmoins impressionnant, lorsqu'on le rapporte au nombre de mutations par gène et par tumeur : dans le cas des tumeurs du sein par exemple, on compte en moyenne 2 mutations *drivers* (Iranzo et al. 2018) et moins de 10 mutations codantes par tumeur (Martincorena & Campbell 2015), soit 10 fois moins. Plusieurs schémas ont été proposés pour l'expliquer, impliquant des mécanismes directs via les DNMTs, recrutées par des facteurs oncogéniques, ou indirects.

Du côté des modifications d'histones (**FIGURE 28**), **l'altération la plus importante dans les cellules cancéreuses est la** réduction globale du niveau de H4K16ac, principalement dans les régions répétées comme les régions sub-téломériques et satellites (Fraga et al. 2005) : cette perte d'acétylation est due à l'activité anormale des HDACs, qui se trouvent être surexprimées ou mutées dans de nombreux types tumoraux (Zhu et al. 2004, Ropero et al. 2006). On note aussi la **perte de H3K4me3** (Fraga et al. 2005) : ces deux modifications pourraient représenter une étape importante de la progression oncogénique, et sont aujourd'hui reconnues comme des caractéristiques typiques des tumeurs, cependant les conséquences fonctionnelles de ces modifications ne sont pas tout à fait élucidées (pour discussion : Henikoff & Shilatifard 2011). Il est possible que **ces altérations déstabilisent la définition et la répression de certaines régions fonctionnelles du génome**, comme les régions satellites ou sub-téломériques. Les cellules cancéreuses affichent également une diminution de la marque activatrice H3K4me3 et des gains des marques répressives H3K9me3 et H3K27me3 dans les régions plus riches en gènes, contribuant à la **répression transcriptionnelle de gènes suppresseurs de tumeurs**, comme *CDKN1A* (Esteller et al. 2007, pour revue : Audia & Campbell 2016). Ces anomalies du patron de modifications des histones sont principalement causées par les expressions aberrantes à la fois des HMTs et des HDMs, mais également par le recrutement fautif de ces enzymes via des modulateurs épigénétiques surexprimés (comme le lincRNA *HOTAIR* dans les cancers du sein, qui interagit avec PRC2 et entraîne un remaniement de la distribution de la marque H3K27me3 – Gupta et al. 2010).

Enfin, **les protéines histones elles-mêmes peuvent subir des mutations oncogéniques** : les plus fréquentes sont les mutations de l'histone H3.3, associées aux sarcomes et aux tumeurs pédiatriques du système nerveux central, et affectant des résidus modifiables post-traductionnellement ou situées à proximité de tels résidus (pour revue : Weinberg et al. 2016). La structure des nucléosomes et leur positionnement peut également être l'objet d'altérations oncogéniques : les remodeleurs chromatiniens BRG1 et BRM par exemple agissent comme des suppresseurs de tumeurs dans les cancers du poumon, et sont réprimés dans 15 à 20% des tumeurs de ce type (Medina et al. 2008). Les variants d'histones sont aussi impliqués dans certains cancers : par exemple, la surexpression de du variant MacroH2A, impliqué en condition physiologique dans la senescence, est associée à un meilleur pronostic dans les tumeurs du poumon (Sporn et al. 2009).

2. L'altération locale des profils épigénétiques modifie l'expression des gènes

Les barrières épigénétiques assurant la stabilité des états de différenciation sont fragilisées lors de la transformation tumorale. Ces aberrations épigénétiques sont influencées par des facteurs environnementaux, lesquels ont bien souvent également un effet génotoxique : on pensera notamment aux inflammations chroniques, à certains carcinogènes comme le tabac, ou au vieillissement en général. Une des conséquences directes de cette fragilisation épigénétique est la dé-répression anormale de centaines de gènes (**FIGURE 29**). En particulier, **l'hypométhylation de l'ADN peut conduire à l'activation de d'oncogènes** (par exemple, *R-RAS* et la cycline D2 dans le cancer de l'estomac, ou *HPV16* dans le cancer du col de l'utérus, [Fainberg & Tycko 2004](#), [Wilson et al. 2007](#)), mais également de gènes spécifiques d'autres tissus (comme les gènes de la famille *MAGEs*, spécifiques du testicule et anormalement exprimés par de nombreux mélanomes). En effet, les cellules cancéreuses expriment un grand nombre de gènes de lignage qui sont censés rester silencieux dans le tissu tumoral d'origine ([Rousseaux et al. 2019](#)), ce qui a attiré l'attention des chercheurs et chercheuses depuis longtemps ; certains ont été particulièrement étudiés : c'est le cas des gènes testis-spécifiques. Réciproquement, **L'hyperméthylation d'ilots CpG en amont de gènes suppresseurs de tumeurs** est également impliquée dans leur perte d'expression (**FIGURE 29**). Elle corrèle positivement avec le degré de malignité des tumeurs : par exemple, le gène suppresseur de tumeur *CDKN2A* est réprimé par méthylation dans respectivement 17% des hyperplasies basales du poumon, 24% des métaplasies, et 50% des carcinomes pulmonaires in situ ([Belonsky et al. 1998](#)). Elle peut aussi participer à l'inactivation totale d'un gène suppresseur de tumeur, en coopérant avec des mutations perte de fonction affectant l'un des deux allèles afin de **satisfaire l'hypothèse « double hits » de Knudson** ([Jones, Issa & Baylin 2016](#)). C'est par exemple le cas dans le cancer du sein, où l'hyperméthylation d'un des allèles de *BRCA1* s'associe fréquemment à la mutation du second allèle pour conduire à la perte de fonction totale de la protéine *BRCA1* : ainsi 20% des tumeurs du sein présentant une mutation perte de fonction de *BRCA1* sont hyperméthylées pour ce gène, contre 5% dans la population de tumeurs non-mutées ([Esteller et al. 2000](#)). Les anomalies de modifications des histones interviennent également dans la répression de certains gènes suppresseurs de tumeurs : par exemple, le gène suppresseur de tumeur *p21^{WAF1}* est réprimé au niveau transcriptionnel dans certaines tumeurs grâce à l'hypoacétylation et l'hyperméthylation des histones H3 et H4 uniquement ([Richon et al. 2000](#)). Enfin, l'affaiblissement des barrières structurales de la chromatine peut suffire à activer certains gènes ([Hnisz et al. 2016](#)).

Réciproquement, la perturbation des voies de signalisation due aux mutations et à l'altération de l'expression des gènes peut modifier le paysage épigénétique. Dans les cancers du sein, la surexpression des récepteurs aux œstrogènes entraîne une activité inappropriée des voies de réponses aux hormones, le recrutement de facteurs pionniers comme *FOXA1* et *GATA3* ainsi que de nombreux co-activateurs et co-répresseurs, occasionnant de profonds remaniements des épigénomes ([Aching-Kawecka et al. 2020](#), revue : [Garcia-Martinez et al. 2021](#)). Ce point est en quelque sorte **le pendant pathologique des mécanismes épigénétiques de propagation des signaux dans le temps** que nous avons évoqués précédemment

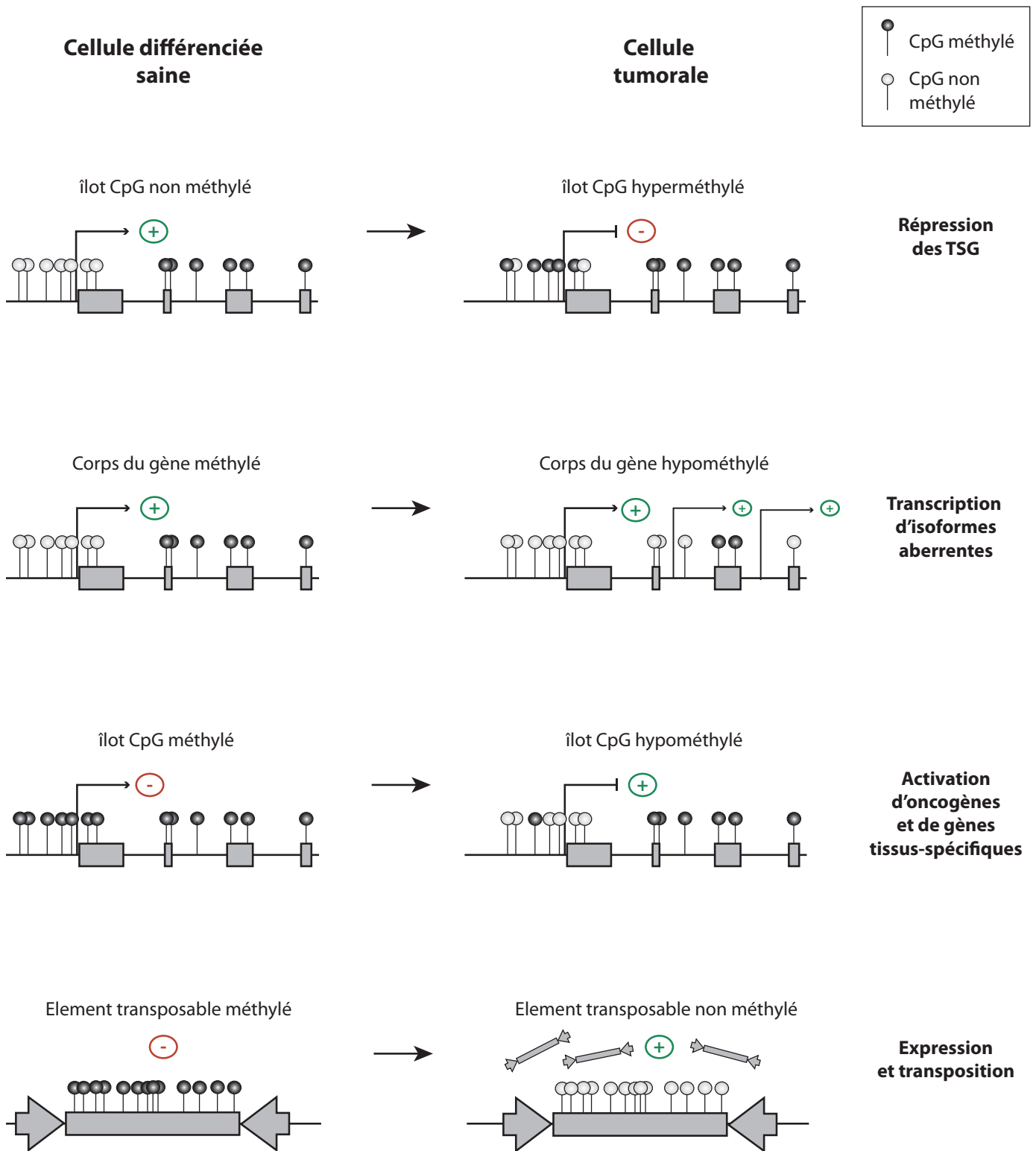


Figure 29 : Anomalies de la méthylation de l'ADN et dérégulation de l'expression des gènes

Quatre situations types, conduisant à des anomalies d'expression des gènes, ont été représentées.

3. La plasticité des états tumoraux est soutenue par des transitions épigénétiques

Au sein d'une tumeur, les interactions réciproques entre cellules tumorales et microenvironnement notamment vont créer une pression de sélection, qui tendra à sélectionner un petit nombre de phénotypes: ce phénomène de sélection des états cellulaires va contraindre des cellules tumorales, éventuellement présentant des mutations différentes à **converger vers des phénotypes similaires**. Réciproquement, grâce notamment aux modulations épigénétiques, des cellules de même génotype peuvent présenter des phénotypes totalement différents. Lors des **transitions épithélio-mésenchymateuses** qui assistent la progression métastatique, les cellules cancéreuses prennent différents états hybrides entre un état cellulaire épithélial, polarisé et solidaire, et un état mésenchymateux, autonome et capable de se déplacer (voir par ex les travaux de Cédric Blanpain : [Pastushenko et al. 2018](#)). Ce spectre phénotypique est établi par le jeu de différents facteurs de transcription et de modifications épigénétiques spécifiques, stabilisant des réseaux de régulation transcriptionnelle différents à génotype égal. Ici, **la plasticité épigénétique soutient et renforce les variations phénotypiques** observées au sein de la tumeur. Plus encore, le dynamisme des modifications épigénétiques va intégrer et amplifier les réponses cellulaires aux pressions environnementales : **les modifications épigénétiques jouent un rôle essentiel dans le comportement des cellules tumorales en réponse aux traitements anti-cancéreux**. Dans le cancer du sein, 60% des tumeurs sont dépendantes de la surexpression de l'oncogène ERα ([Yager & Davidson 2006](#)) : le traitement standard de ces tumeurs inclue donc des thérapies endocrines. Cependant, un tiers de ces patientes développeront une résistance au traitement et verront leur maladie progresser sous thérapie endocrine ([Haque & Desai 2019](#)). Certaines études y ont démontré le rôle de l'épigénétique : le niveau d'expression de ER ainsi que de ses gènes cibles peut être modulé par les cellules cancéreuses grâce à l'hyperméthylation du gène *ESR1* ([Yang et al. 2001](#) ; [Kangaspeska et al. 2008](#)), un mécanisme pouvant expliquer l'émergence de clones résistant aux hormonothérapies. Viser les régulateurs épigénétiques afin de modifier l'épigénome des cellules cancéreuses pourraient permettre de re-sensibiliser la tumeur aux thérapies endocrines.

Les altérations épigénétiques des cellules cancéreuses sont une modification des fonctions chromatiniennes plus « douce » que les altérations de la séquence d'ADN : en effet les marques épigénétiques sont potentiellement réversibles. Cette constatation a entraîné une réflexion autour du développement de « **thérapies épigénétiques** » ([Allis & Jenuwein 2016](#)). Grâce à l'intervention de traitements pharmacologiques inhibant l'activité de certains modulateurs épigénétiques (par exemple, des inhibiteurs de HDACs ou des DNMTs), on pourrait alors modifier le phénotype des cellules cancéreuses et améliorer l'évolution de la maladie. Cet axe de recherche est aujourd'hui très prolifique, et on assiste au développement de multiples études pré-cliniques et cliniques impliquant ce type de thérapies épigénétiques. Ces approches s'appuient évidemment sur l'identification de cibles épigénétiques spécifiques des cellules tumorales : il est donc nécessaire de caractériser les altérations épigénétiques fréquemment observées dans les tumeurs et leur rôle dans la progression tumorale. C'est avec cet objectif que je me suis intéressée à la régulation épigénétique des gènes Cancer/Testis dans les tumeurs.

Partie III

**Les gènes Cancer/Testis sont-ils un modèle
crédible pour investiguer les altérations
épigénétiques des cellules cancéreuses ?**

1. Des gènes épigénétiquement réservés à la lignée germinale mâle...

L'identification des gènes testis-spécifiques a progressé parallèlement avec l'évolution des techniques de séquençage, permettant l'établissement des profils d'expression précis des gènes du génome humain à travers les différents tissus. Ainsi, plusieurs listes définissant les gènes testis-spécifiques ont pu être établies au cours des dernières années (Messmer et al. 2009 ; Rousseaux et al. 2013 ; Wang et al. 2015), identifiant un nombre toujours plus important de gènes présentant cette spécificité d'expression. Cependant, ces listes ne se recoupent qu'imparfaitement : en effet, outre l'évolution des techniques de séquençage, il n'existe pas aujourd'hui de consensus clair concernant les méthodes d'analyse et d'identification des gènes tissus-spécifiques. Aujourd'hui, on compte plus d'une quarantaine de familles représentant plus de 1000 gènes, tous présentant la particularité d'avoir une expression restreinte au testis. Notons que certains tissus somatiques expriment quelques gènes C/T mais, d'après des données de RT-qPCR, le niveau d'ARNm des gènes C/T dans les tissus somatiques est généralement inférieur à 1% de leur expression par les gamètes mâles (Caballero & Chan 2009, Scanlan et al. 2004).

Des études de KO ont mis en lumière la diversité des fonctions exercées par ces protéines. On observe fréquemment une altération de la fertilité chez des modèles murins déficients pour ces gènes testis-spécifiques. Tous ne sont pas exprimés aux mêmes étapes de la gamétogenèse : certains interviennent dans les étapes très précoces de la spermatogenèse, comme MAGE-A1 et NY-ESO-1, alors que d'autres sont spécifiques des spermatozoïdes matures (**FIGURE 30**). Parmi les fonctions identifiées, on citera : les fonctions relatives à la **méiose** (avec en particulier la transesterase SPO11, et les protéines du complexe synaptonémal SYCE1, SYCP1 et HORMAD1) ; le maintien de l'**intégrité du génome** suite aux remaniements épigénétiques de la spermatogenèse (et notamment les protéines PIWIL2 et TDRD1, qui s'associent aux piRNA et semblent importantes pour maintenir la répression des transposons durant la spermatogenèse, mais aussi TDRD6 et MAEL) ; la **régulation transcriptionnelle** des gènes spécifiques des gamètes mâles, où interviennent CTCFL (également appelé BORIS, un *master regulator* d'une partie du programme transcriptionnel spécifique des gamètes (Suzuki et al. 2010), BRDT (une protéine à bromodomaine essentielle à la différenciation des gamètes mâles (Shang et al. 2007) et ATAD2 (une autre protéine à bromodomaine agissant comme cofacteur de l'oncogène MYC) ; la **mobilité des spermatozoïdes**, avec les protéines architecturales du flagelle (AKAP4, CABYR, SPA17, ROPN1...) ou de l'acrosome (ACRBP, DKKL1...), et les protéines nécessaires à la fertilisation (CALR3, ADAM2...) ou à l'implantation comme PLAC1 ; enfin le **métabolisme** particulier des spermatozoïdes, avec notamment les protéines mitochondriales testis-spécifiques SPATA19 et COX6B2, ou encore la lactate déshydrogénase C LDHC.

Cependant, de nombreux gènes testis-spécifiques n'ont pas encore révélé leur fonction : pour certains, ce sont des gènes n'ayant pas d'homologue chez les rongeurs (comme le gène spécifique des spermatozoïdes matures CT83), ce qui rend leur étude plus difficile.

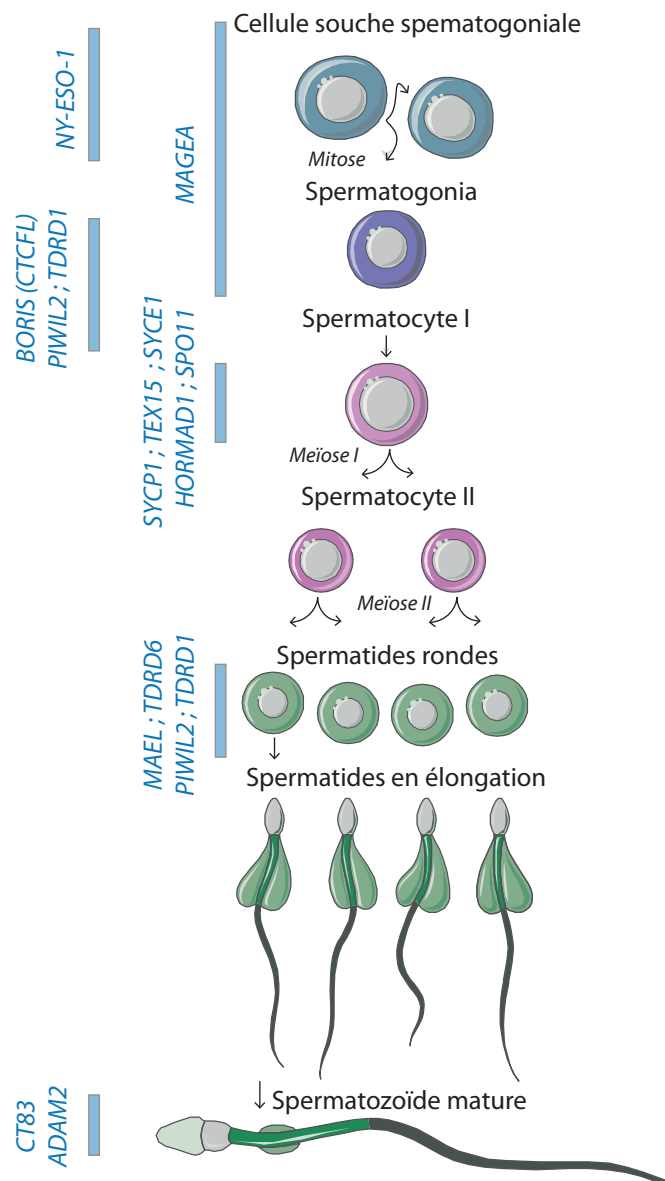


Figure 30 : Expression des gènes C/T au cours de la différenciation des gamètes mâles

L'expression des gènes C/T est précisément régulée au cours de la différenciation des spermatozoïdes : certains gènes C/T sont spécifiques de stades particuliers au cours de la différenciation. Quelques gènes C/T spécifiques ont été soulignés, il en sera question au cours de notre exposé.

2. Des gènes anormalement exprimés par de nombreuses tumeurs

De telles fonctions (motilité, métabolisme anaérobie, prolifération...) peuvent être détournées à l'avantage des cellules cancéreuses. De fait, tous les gènes testis-spécifiques que nous avons cité plus haut se trouvent être exprimés par un ou plusieurs types de tumeurs ; plus encore, de tous les gènes tissus-spécifiques anormalement exprimés par les cellules cancéreuses, les gènes testis-spécifiques sont les plus fréquemment activés (**FIGURE 31**, Rousseaux et al. 2019).

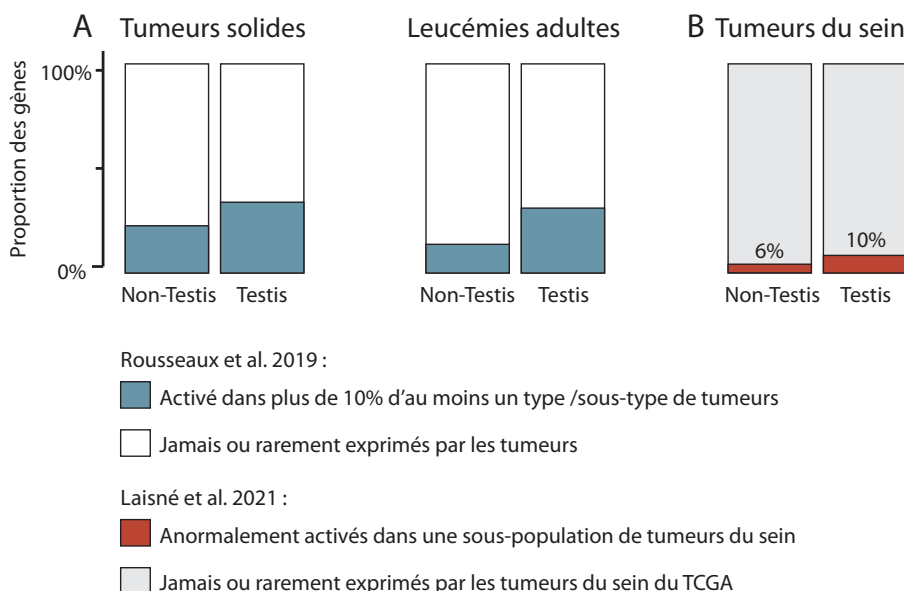


Figure 31 : Proportions de gènes tissus-restreints anormalement activés par les tumeurs, en fonction de leur origine

Les gènes tissus restreints ont été définis d'après une analyse maison (A) ou extrapolés à partir de Kim et al., 2017 (B). Les données d'expression sur les tumeurs proviennent de trois séries publiques : GSE2109 (étude multicancers), GSE13159 (leucémies adultes), et TCGA BRCA (cancers du sein). Les critères retenus pour définir une expression atypique sont plus ou moins stringents, mais on retrouve le même résultat : proportionnellement deux fois plus de gènes tissu-restreints exprimés par les tumeurs sont d'origine testiculaire. (Adapté de Rousseaux et al. 2019, Laisné et al. 2021)

Certaines tumeurs (mélanomes, vessie, poumon) les activent très fréquemment (Scanlan et al. 2004), au contraire de tumeurs (rein, colon) qui les expriment rarement. **Les cancers du sein présentent un profil d'expression des gènes C/T intermédiaire.** Dans un même type tumoral, les différents gènes C/T ont des probabilités d'activation différentes : les plus fréquemment activés sont NY-ESO-1 (17 à 42% des mélanomes l'expriment) et MAGEA3 (57 à 76% des mélanomes). **Les gènes C/T ont tendance à être co-exprimés** : en 1998 Sohin a montré que si 47% des tumeurs du sein n'exprimaient aucun des gènes C/T étudiés, 40% en exprimait au moins trois (Sphin et al. 1998). Certains gènes C/T ont leur loci disposés en cluster (MAGEAs, MAGEBs, MAGECs...) : on observe alors fréquemment une co-activation de l'ensemble des gènes C/T du cluster. Enfin, l'analyse de coupes de tumeurs par immunohistochimie a permis d'évaluer la distribution intratumorale de l'expression des gènes C/T, grâce à des anticorps monoclonaux (Jungbluth et al. 2000, Jungbluth et al. 2001, dos Santos et al. 2000). Ces analyses ont

révélé la **grande variabilité d'expression des gènes C/T au sein d'une même tumeur** : dans le cas de carcinomes ductal in situ du sein par exemple, l'expression de NY-ESO-A pouvait être selon les tumeurs intense et largement répandue (plus de 90% des cellules tumorales détectées comme positives), ou au contraire sporadique, avec une minorité (>50%) de cellules positives regroupées en îlots focaux, ou encore dispersées à travers la masse tumorale (Caballero et al. 2014).

Une proportion significative de gènes C/T possède un promoteur doté d'un îlot CpG (FIGURE 32) : en effet, et parce que ces îlots CpG sont majoritairement exprimés et déméthylés dans les cellules germinales, ils se trouvent épargnés par la déamination mutagène qui affecte les CpG méthylés et les fait progressivement disparaître . Pour ces gènes, l'hypométhylation de ces îlots est fréquemment corrélée avec leur activation dans les tumeurs qui les expriment. Néanmoins tous les gènes C/T n'ont pas d'îlots CpG et ce mécanisme n'est pas une règle absolue. Pour conclure, on observe une association préférentielle entre le type tumoral et le pool de gènes C/T activés : ce résultat suggère que certains facteurs pourraient favoriser l'activation des gènes C/T. Les éléments de contexte cellulaire accompagnant la transformation pourraient alors moduler la probabilité d'activation de certains gènes C/T.

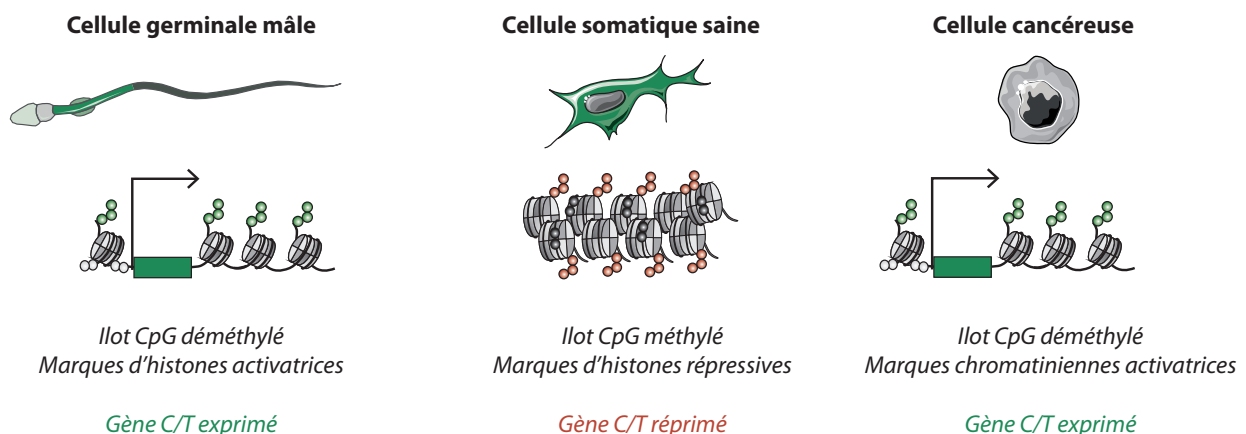


Figure 32 : Expression des gènes C/T

Pour cette raison, l'étude de l'expression des gènes C/T constitue un modèle de choix pour explorer la question fondamentale des interactions entre signalisation et régulation épigénétique dans l'établissement des programmes de différenciation cellulaire. D'autre part, l'étude de leur expression hors-contexte dans les cancers offre un angle d'approche intéressant pour analyser les anomalies de régulations transcriptionnelles accompagnant l'oncogenèse. De plus, l'expression des gènes C/T par les tumeurs offre des perspectives thérapeutiques intéressantes pour le traitement des tumeurs.

3. Les gènes C/T sont un objet d'étude prometteur en oncologie

Historiquement, les gènes C/T ont été étudiés comme une cible potentielle d'immunothérapie. L'intérêt pour ces gènes s'est développé dans les années 1940-1950, où une série d'expériences ont montré que **certaines tumeurs exprimaient des facteurs immunogènes, reconnus comme étrangers et pouvant provoquer leur élimination par le système immunitaire d'une souris syngénique** (à la manière du rejet d'un organe transplanté). Ces antigènes reconnus comme cibles par le système immunitaire sont de différentes natures : ce peuvent être des antigènes provenant d'oncovirus comme E6 et E7 ; des néoantigènes provenant de la dégradation de protéines mutées (comme RAS ou TP53) ou de protéines chimériques issues de réarrangements géniques (comme les antigènes issus de la protéine de fusion BCR-Abl), des antigènes présents en une quantité anormale suite à la surexpression de leur protéine d'origine (comme HER2), mais également des antigènes issus des gènes C/T. En effet, de part leur spécificité d'expression, les antigènes C/T exprimés sont protégés par la barrière hémato-testiculaire, et donc inaccessibles aux cellules immunitaires de la circulation générale. De ce fait, ces antigènes sont moins susceptibles de faire l'objet d'une tolérance périphérique que les antigènes spécifiques des autres tissus. Ces protéines exprimées par les cellules cancéreuses pourraient être à l'origine d'antigènes immunogéniques détectés par les cellules immunitaires, provoquant une réponse immunitaire anti-cancéreuse. De surcroît, le rôle pro-oncogénique de certains gènes C/T dans le contexte tumoral a été démontré : l'activation illégitime des gènes C/T participe à l'acquisition des caractéristiques fondamentales des cancers (**FIGURE 33**).

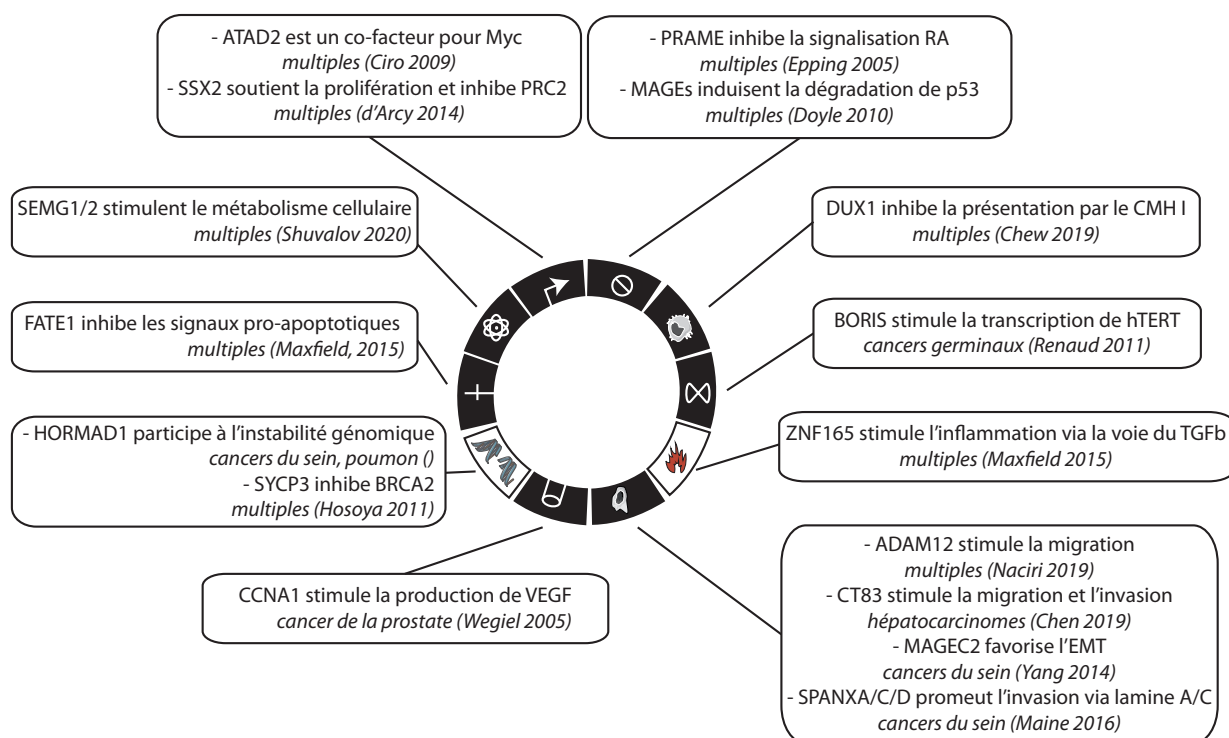


Figure 33 : Les gènes C/T peuvent agir en tant qu'oncogènes

Synthèse d'une revue de littérature reprenant différents rôles des gènes C/T dans la tumorigenèse de certains cancers, en lien avec les caractéristiques des cancers telles que définies par Weinbergs.

Le premier antigène C/T a été découvert au début des années 1990 chez une patiente atteinte d'un mélanome dont l'évolution était étonnamment favorable : chez cette patiente, les cellules cancéreuses du mélanome exprimaient le gène **MAGE-A1** (*Melanoma Antigen Gene A1*), normalement absent des mélanocytes et reconnu par les lymphocytes T cytotoxiques (van der Bruggen et al. 1991). D'autres tumeurs expriment également *MAGE-A1* (notamment des carcinomes du sein), mais cette expression est indétectable dans les tissus sains à l'exception du testicule (Chen et al. 1994). Cette découverte a rapidement été suivie de l'identification d'autres facteurs présentant les mêmes propriétés d'expression que *MAGE-A1*, tels que les gènes C/T des familles BAGE et GAGE. La découverte des gènes C/T a largement bénéficié de l'approche SEREX (*serological identification of antigens by recombinant expression cloning*) (Sahin et al. 1995), qui part de l'analyse des anticorps présents dans le sérum des patients atteints de cancer pour identifier des antigènes. Le facteur *NY-ESO-1* (*New York esophageal squamous cell carcinoma*) a ainsi été identifié par cette méthode et est aujourd'hui reconnu comme l'un des antigènes C/T les plus immunogènes (Chen et al. 1997). Ces dernières années, **les approches onco-immunologiques des gènes C/T** ont connu un certain succès, comme en témoigne le lancement en 2007 d'une étude de phase III portant sur un vaccin ciblant *MAGE-A3* pour les mélanomes et les cancers du poumon (GSK, Brichard et al. 2008). Une autre approche prometteuse, cette fois-ci pour le glioblastome, utilise les lymphocytes T helper CD4+ des patients afin de générer une réponse immunitaire dirigée contre les gènes C/T exprimés par les cellules tumorales : dans ce protocole, les lymphocytes T sont isolés et traités avec un agent déméthylant, afin de leur permettre d'exprimer et de présenter de nombreux gènes C/T. Ils sont ensuite directement utilisés comme cellule présentatrice d'antigènes afin de stimuler une réponse cytotoxique et NK chez ces patients. Dans un essai de phase I, cette approche a été utilisée pour 25 patients atteints de glioblastomes réfractaires au traitement standard, et 3 d'entre eux ont montré une réponse durable avec une régression significative de la masse tumorale (Kirkin et al. 2018). Cette approche, par **transfert adoptif de lymphocytes T stimulés ex vivo**, a également montré des résultats très intéressants dans le cancer du col de l'utérus positif pour le papillomavirus (HPV) : contre toute attente, l'une des patientes présentant une réponse complète sous cette immunothérapie affichait une majorité de clones T reconnaissant non pas HPV, mais un gène C/T anormalement exprimé par les cellules cancéreuses du col de l'utérus, CT83 (Stevanovic et al. 2017).

Au cours de ma thèse, je me suis intéressée à la régulation des gènes Cancer/Testis comme modèle pour étudier les modifications épigénétiques. J'ai pu développer deux axes complémentaires, en investiguant d'une part les mécanismes moléculaires à l'origine de leur répression dans les cellules non transformées ; et d'autre part les mécanismes et le rôle de leur activation anormale par les cellules cancéreuses. Dans ce second projet, je me suis concentrée sur les cancers du sein : en effet, l'expression atypique d'un marqueur des cellules germinales mâles par les cellules cancéreuses de la glande mammaire offre tout particulièrement de nombreuses applications cliniques, et notamment dans le développement de traitements immuno-oncologiques visant ces cibles très spécifiques et sûres. C'est pourquoi je finirai cette introduction par une présentation générale de mon modèle d'étude, les cancers du sein.

CHAPITRE 3

Le cas des cancers du sein

Partie I

Quelle est l'origine des cancers du sein ?

1. Epidémiologie et facteurs de risque :

Le **cancer du sein** est de loin **le cancer le plus fréquent chez les femmes** avec 1,7 millions de cas diagnostiqués dans le monde (ce qui correspond à 25% de l'ensemble des diagnostics de cancers) et plus de 500 000 décès estimés pour 2012 (Ferley et al. 2015). Ces données placent le cancer du sein en 5^e position des causes de décès par cancer, à l'échelle mondiale. En France, les données actuelles estiment **qu'une femme sur neuf** sera concernée au cours de sa vie, avec environ 54 000 nouveaux cas et 12 000 décès par an (Binder Foucard et al. 2013). L'incidence de ce cancer a beaucoup augmentée entre 1980 et 2000, et est en légère diminution depuis 2005 : plusieurs facteurs pourraient expliquer cette baisse, notamment la saturation du dépistage organisé ou la baisse de la prescription des traitements hormonaux substitutifs de la ménopause (rapport INCA 2017). Grâce aux progrès de la prise en charge thérapeutique, le taux de mortalité du cancer du sein diminue depuis les années 80 (FIGURE 34). Le taux de survie à 5 ans après le diagnostic est de 87%, celui à 10 ans de 76%. Malgré ces résultats encourageants, le cancer du sein reste la **première cause de décès par cancer chez les femmes en France** (18% des décès féminins par cancer), et demeure un enjeu majeur de santé publique (Ribassin-

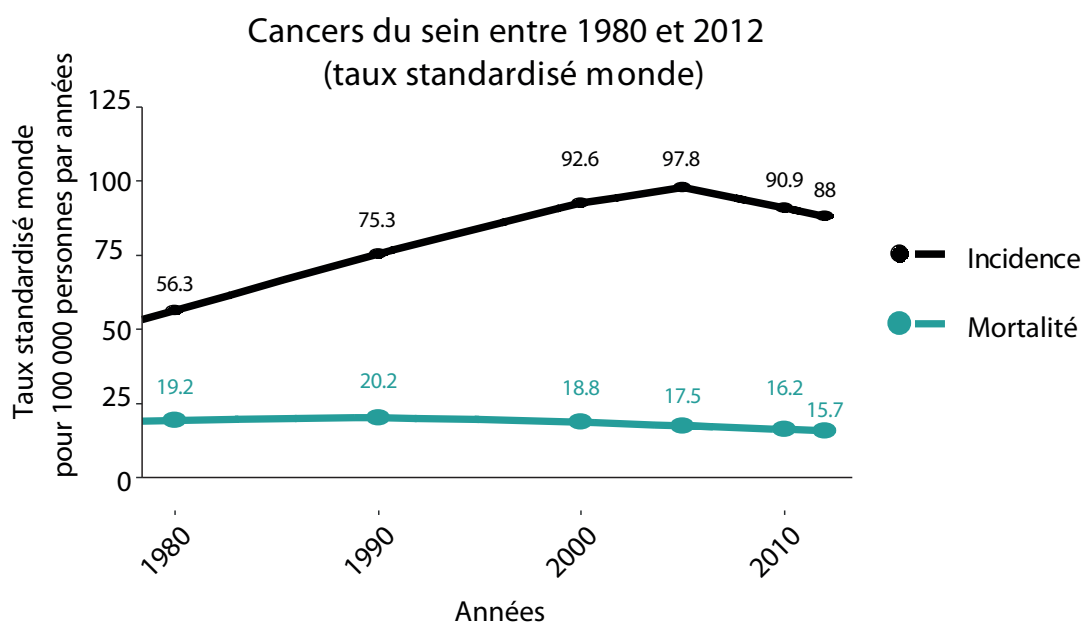


Figure 34 : Evolution des cancers du sein entre 1980 et 2012

L'évolution de l'incidence des cancers du sein dans le monde (taux standardisé estimé) montre une première inflexion en 2005, et tend à diminuer depuis. La mortalité diminue régulièrement depuis les années 80 (en moyenne 1,5% par an). (Source : INCA 2013, Bonder-Foucard 2013)

Majed et al. 2017).

Plusieurs facteurs de risque spécifiques aux cancers du sein sont connus : l'âge tout d'abord (l'incidence augmente après 30 ans, et l'âge médian au diagnostic est de 63 ans), les facteurs hormonaux (le risque de cancer du sein augmente avec la durée d'exposition aux œstrogènes : ainsi un âge précoce à la ménarche, un âge tardif à la ménopause, l'absence de grossesse sont des facteurs de risque), les antécédents d'hyperplasie atypique mammaire, les antécédents de radiothérapie thoracique chez l'enfant ou l'adolescent, et l'obésité. D'autres facteurs de risques comportementaux ont été identifiés

comme les facteurs alimentaires (dont la consommation d'alcool), tandis que l'activité physique a clairement été établie comme un facteur protecteur. Soulignons que les cancers du sein existent aussi chez les hommes, même s'ils restent exceptionnels (moins de 1% des cancers du sein). Les antécédents familiaux influencent le risque individuel de développer un cancer du sein. Plusieurs mutations germinales de prédisposition sont aujourd'hui connues : en particulier les **mutations des gènes BRCA1 et BRCA2** (*BReast Cancer 1, 2*), qui ont une prévalence de 2-3% aux Etats-Unis (Winters et al. 2017). *BRCA1* et *BRCA2* codent pour des protéines qui participent au maintien de l'intégrité du génome. Ce sont des gènes suppresseurs de tumeurs essentiels : la présence d'une mutation sur l'un de ces deux gènes fait passer le risque d'apparition d'un cancer du sein à 50-80% chez les porteuses de ces mutations (contre 10% dans la population générale), avec un âge médian au diagnostic significativement plus jeune (45 ans Kuchenbaecker et al. 2017 ; Howlader et al. 2020). Ces variants génétiques sont donc considérés comme des mutations rares dans la population, mais avec une pénétrance très élevée.

2. Architecture de la glande mammaire

D'un point de vue anatomique, **le sein se compose d'une glande mammaire entourée par un stroma** adipeux, conjonctif et richement vascularisée (FIGURE 35). La glande mammaire s'organise en un réseau de canaux épithéliaux formant 15 à 20 lobes, eux-mêmes constitués de lobules capables de sécréter du lait en période d'allaitement. Les lobes sont drainés par des canaux galactophores qui convergent vers le mamelon (Winslow et al. 2011). La morphogenèse de la glande mammaire a principalement lieu durant la période post-natale, et cet organe hautement spécialisé connaîtra des épisodes d'expansion, de différenciation et de mort cellulaire à chaque cycle de reproduction (Fu et al. 2020).

Le tissu épithélial mammaire est composé principalement de deux types cellulaires (FIGURE 35-36) : les cellules épithéliales, dites luminales, qui définissent le lumen central des canaux, entourées par une couche externe de cellules myoépithéliales (dites également basales) directement en contact avec la membrane basale qui sépare les canaux et les acini de la membrane extracellulaire environnante. La fonction principale des cellules luminales est de produire le lait via les cellules sécrétrices, tandis que les propriétés contractiles des cellules myoépithéliales permettent l'expulsion du lait. Ces deux types cellulaires constituent ensemble un système de canaux à deux couches, qui évoluera tout au long de la vie des mammifères (Macias & Hinck, 2012).

Pour assurer ces changements morphologiques, **la glande mammaire abrite des cellules souches adultes.** Une cellule souche adulte est définie comme une cellule non-spécialisée possédant des propriétés d'auto-renouvellement quasi-indéfini et de différenciation multipotente afin de générer tous les types cellulaires composant un tissu ou un organe donné (Visvader et al. 2016). Les **cellules progénitrices**, au contraire, présentent des propriétés de différenciation plus limitées et ont perdu la capacité d'auto-renouvellement des cellules souches. **Les cellules souches adultes de la glande mammaire ont été identifiées formellement par deux types d'expérience.**

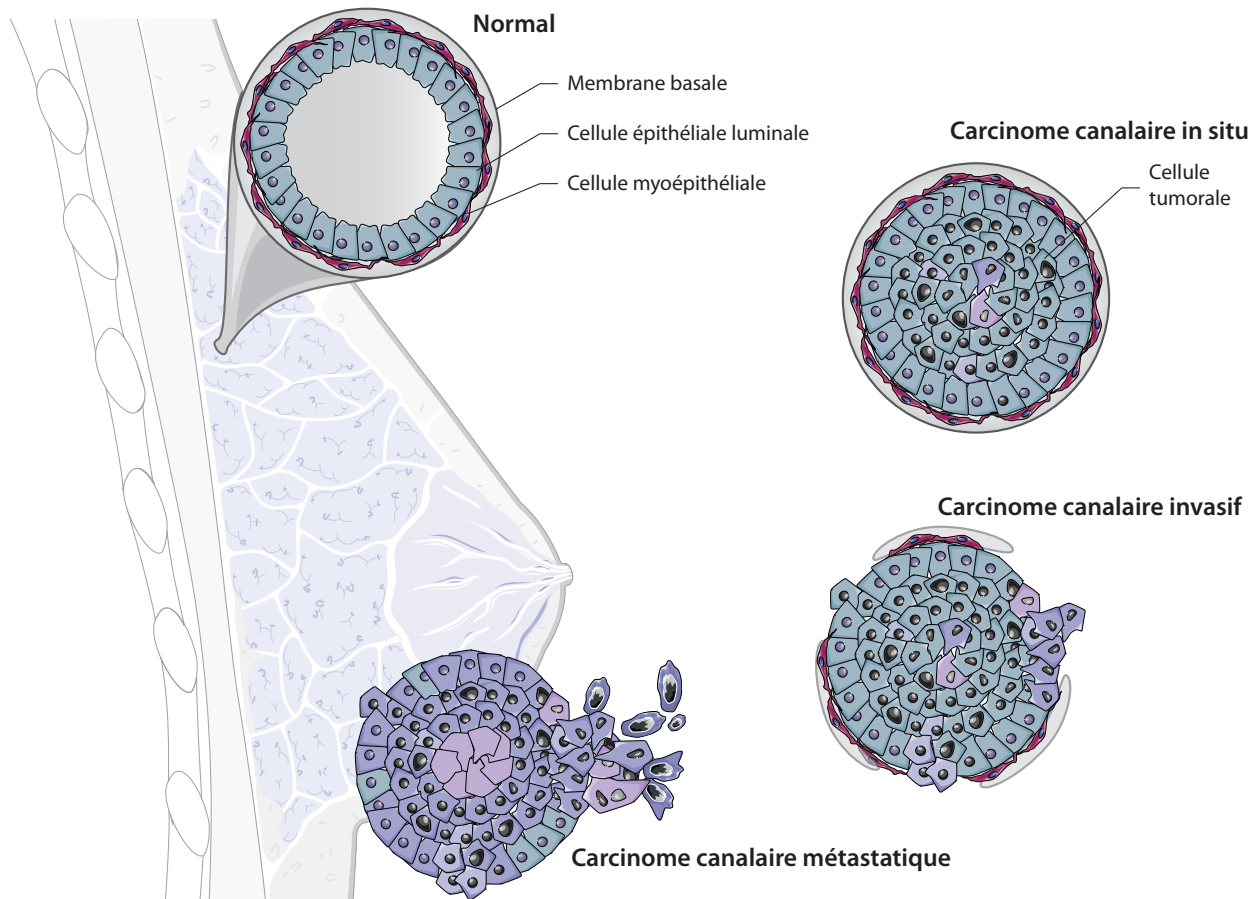


Figure 35 : Classification anatomopathologique des cancers canaux

Les anomalies tissulaires causées par le développement d'un carcinome canalaire peuvent rester confinées au sein de la membrane basale, qui demeure intacte, et envahir uniquement la lumière des canaux. Certaines tumeurs évolueront jusqu'à franchir la membrane basale et envahir le compartiment stromal, ce qui leur permet d'infiltrer les vaisseaux sanguins et les canaux lymphatiques environnants, puis éventuellement de métastaser. (Adapté de Marshall 2014)

D'une part, les expériences de transplantation ont exploré la capacité inhérente de cellules isolées depuis une glande mammaire différenciée à régénérer une glande mammaire fonctionnelle dans un nouvel environnement vierge (un *fat pad* duquel on a chirurgicalement enlevé la glande mammaire elle-même). Par définition, seules les cellules multipotentes en sont capables (voir par exemple [Prater et al. 2014](#)). D'autre part, les expériences de *lineage* tracing suivront au cours du temps le destin cellulaire des cellules souches et des progéniteurs dans leur environnement d'origine. Ces deux types d'études ont permis de proposer un modèle hiérarchique de différenciation pour les cellules épithéliales de la glande mammaire (**FIGURE 36**). Même si ce modèle semble unidirectionnel, certaines études suggèrent que des phénomènes de trans-différenciation peuvent avoir lieu, en particulier lors de la néoplasie. **La dissection précise de l'hétérogénéité cellulaire composant la glande mammaire normale est importante pour comprendre l'existence de différents sous-types de cancers du sein.**

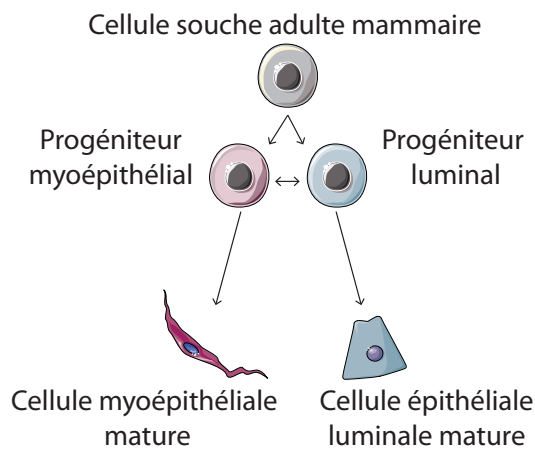


Figure 36 : Sous-types de cellules épithéliales de la glande mammaire normale

3. Processus de transformation carcinomateuse

L'oncogenèse des carcinomes du sein est généralement initiée par une ou plusieurs mutations transformant une cellule épithéliale saine. **Un stade préliminaire à la cancérisation est défini par l'hyperplasie**, qui décrit un état de prolifération excessive des cellules épithéliales, produisant une accumulation de cellules ayant une morphologie normale. On parlera **d'hyperplasie atypique** lorsque les cellules qui prolifèrent prennent certaines caractéristiques anormales : on verra apparaître des altérations morphologiques, une intensification du caractère prolifératif et l'acquisition d'un caractère plus indifférencié, qui témoignent d'une perte partielle ou totale de leur fonction dans l'organisme.

Les cancers du sein se développent à partir des canaux (cancers canaux) ou des lobules (cancers lobulaires) de la glande mammaire. **La majorité des cancers du sein sont d'origine canalaire**. On distingue deux grandes catégories (**FIGURE 35**) : **les cancers *in situ*, qui représentent environ 20% des cancers du sein** et dont 85% à 90% sont des carcinomes de type canalaire, lorsque les cellules cancéreuses restent confinées dans la lumière des canaux et lobules. **Les cancers sont dits infiltrants ou invasifs (75% des cas)**, lorsque les cellules cancéreuses traversent la membrane basale et colonisent le tissu avoisinant. Les cellules malignes peuvent alors éventuellement se propager dans l'organisme via les vaisseaux sanguins et lymphatiques et former des métastases. Les principaux organes touchés par les métastases issues des cancers du sein sont le foie, les os et les poumons ([Lakhani et al. 2012](#)).

De nombreuses données intégratives suggèrent que les lésions des cancers canaux *in situ* seraient des précurseurs des lésions de carcinomes canaux infiltrants, sans qu'elles soient nécessairement obligatoires. **Bien que tous les cancers *in situ* ne progressent pas vers un cancer invasif, tout l'enjeu de leur prise en charge réside dans le risque de récurrence invasive** (50% des récurrences surviennent sous forme invasive), qui peuvent être très tardive (jusqu'à 40 ans après le diagnostic de carcinome canalaire *in situ* en l'absence de traitement local) ([Sanders et al. 2005](#)).

Partie II

Comment classer les cancers du sein ?

1. Classification histologique

Afin d'évaluer le développement d'une tumeur, le corps médical utilise la classification TNM (FIGURE 36), qui prend en compte trois éléments : la taille de la tumeur T (*tumor*), l'envahissement ganglionnaire N (*node*) et la présence de métastases à distance M (*metastasis*). **Ces trois éléments permet de définir des stades**, qui apportent une indication pronostique. : en effet, la survie relative à 5 ans passe de 99% pour un cancer du sein localisé, à 85% en cas de cancer avec envahissement ganglionnaire régional, pour diminuer à 27% en cas de cancer métastatique (Howlader et al. 2019). **L'analyse post-chirurgicale permet également de définir un grade histologique (FIGURE 36)**. Trois paramètres morphologiques sont évalués : la différenciation architecturale de la tumeur, le pléomorphisme nucléaire, et l'activité mitotique. La somme des trois évaluations correspond au grade et est corrélé au degré de malignité de la tumeur. Comme la classification TNM, la définition du grade a valeur pronostique, que ce soit pour les groupes de patient avec ou sans envahissement ganglionnaire régional. En effet, la survie relative à 10 ans passe de plus de 90% pour les tumeurs de grade I à 70% environ pour les tumeurs de grade III.

2. Classification sur la base de l'expression des récepteurs hormonaux et de HER2

Historiquement, trois marqueurs histologiques ont été identifiés permettant une première classification des tumeurs du sein : ce sont **la surexpression des récepteurs hormonaux aux estrogènes (ER) et à la progestérone (PR), et la surexpression/amplification de l'oncogène ERBB2 (FIGURE 37)**. Ces éléments permettent également d'affiner le pronostic des patientes, et de choisir une stratégie thérapeutique

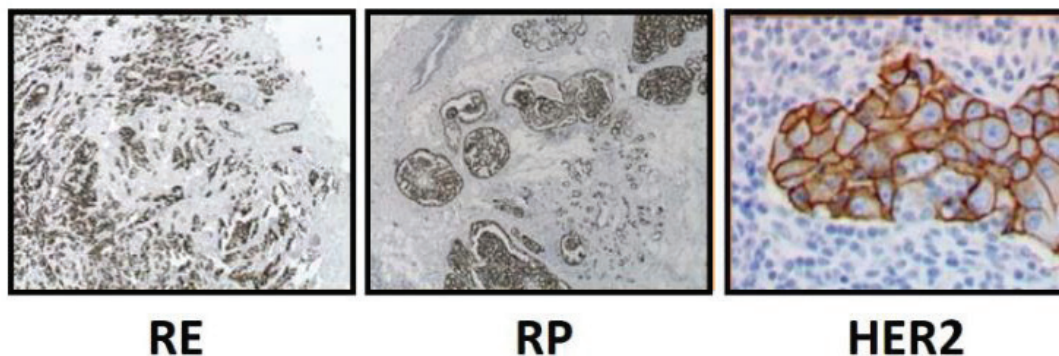


Figure 37 : Exemples de marquage immunohistochimiques (IHC)

De gauche à droite : IHC de tumeurs positives pour, respectivement : les récepteurs aux estrogènes (RE), à la progestérone (PR), et le récepteur HER2. Source : Dr. Anne Vincent-Salomon, Institut Curie

ER et PR sont des facteurs de transcriptions, qui lorsqu'ils sont activés stimulent la croissance des cellules de l'épithélium mammaire sain, mais également la prolifération des cellules cancéreuses les surexprimant. La surexpression de ER et PR a généralement une origine transcriptionnelle. **Ces cancers sont dit hormono-dépendants, représentent la majorité des cancers du sein (70-80%) et offrent le meilleur pronostic.** L'hormonothérapie est le traitement clé de ces cancers ; les estrogènes étant incriminés dans le développement et la récurrence de ces tumeurs du sein. L'utilisation du tamoxifène,

premier anti-estrogène disponible, a ainsi révolutionné le traitement hormonal de ces cancers qui, auparavant, reposait essentiellement sur la suppression ovarienne par radiothérapie ou chirurgie (Obiorah et al. 2011).

ERBB2 est un oncogène de la famille des récepteurs de facteurs de croissance à activité tyrosine kinase (RTK), pouvant dimériser après la fixation de l'EGF et activer les voies de signalisation des MAP kinases et PI3K/AKT. **Les tumeurs dites HER2+ présentent une forte expression de HER2, généralement due à l'amplification du gène ERBB2** situé sur le chromosome 17. La concentration en protéine HER2 à la surface des cellules entraîne l'activation prolongée des voies de signalisation en aval, impliquées dans la survie cellulaire et la prolifération. La détermination du statut HER2 portait initialement sur une évaluation pronostique (**15-20% des cancers du sein surexpriment HER2, initialement de mauvais pronostic**), puis est devenue systématique avec l'avènement des thérapies ciblées anti-HER2 afin de poser leur indication. Pour ces cancers, un anticorps conçu pour inhiber spécifiquement les récepteurs HER2 a été développé et constitue un exemple modèle de thérapie ciblée., qui s'est traduit en clinique par une amélioration remarquable du pronostic des patientes (Piccart-Geghart et al. 2005).

Enfin, le statut **Triple-Négatif** définit des tumeurs négatives pour les deux récepteurs hormonaux et pour le statut HER2. **Ces tumeurs représentent 15 à 20% des cancers du sein, et sont de mauvais pronostic**, avec un caractère très agressif et métastatique. Le traitement de ces cancers n'a pas connu d'évolution majeure depuis l'avènement des chimiothérapies : en effet, et contrairement aux trois autres groupes de cancers du sein qui peuvent bénéficier d'une hormonothérapie ou d'un traitement anti-HER2, ces tumeurs ER-/PR-/HER2- ne bénéficient actuellement d'aucune thérapie ciblée (à l'exception de certaines tumeurs porteuses de mutations spécifiques - BRAC1/2 - et éligibles à des traitements par inhibiteurs de PARPs). L'identification de nouvelles solutions thérapeutiques et de nouvelles cibles manipulables par la pharmacopée conventionnelle constitue donc un enjeu capital pour ces cancers.

3. Classifications moléculaires

Plus récemment, le développement de techniques de séquençage à haut débit comme les puces à ADN a permis d'affiner la classification des cancers du sein, grâce à l'analyse simultanée de plusieurs milliers de gènes. Ces travaux ont été initiés par Perou & Sørlie (Perou et al. 2000, Sørlie et al. 2001) et ont été prolongés depuis (Sørlie et al. 2003, Rakha et al. 2008, Reddy et al. 2011). Les travaux de Perou & Sørlie ont défini **au moins 5 sous-types moléculaires**, à partir du clustering hiérarchique de tumeurs du sein sur l'expression de 1700 gènes (FIGURE 38A). Ces analyses ont confirmé le comportement biologique différent des tumeurs ER+/PR+ par rapport aux autres tumeurs : les tumeurs ER+/PR+ appartiennent principalement aux groupes Luminal A et B, et expriment de nombreux gènes caractéristiques des cellules luminales mammaires (dont *ESR1*) ; alors que les tumeurs ER-/PR- composent principalement le sous-type *basal-like* et expriment des gènes spécifiques des cellules myoépithéliales (comme les gènes des cytokératines 5, 14 et 17). Les cancers *basal-like* ont le plus mauvais pronostic (FIGURE 38B). Cette classification moléculaire a également **révélé deux catégories de tumeurs ER+/PR+** : les tumeurs luminal B sont caractérisées par une expression plus importante de gènes impliqués dans la prolifération (notamment Ki-67), et une plus faible expression de gènes spécifiques de la lignée luminaire comme PR et FOXA1. En adéquation avec leur comportement plus prolifératif, les tumeurs luminal B sont de moins bon pronostic que les tumeurs luminal A (FIGURE 39). Ces classifications ont ensuite évolué vers des tests plus compatibles avec la routine clinique, basés sur la technique de PCR et mesurant un petit nombre de gènes (par exemple la signature *PAM50* ou *Prosigna*®, basée sur l'évaluation d'une signature de 50 gènes).

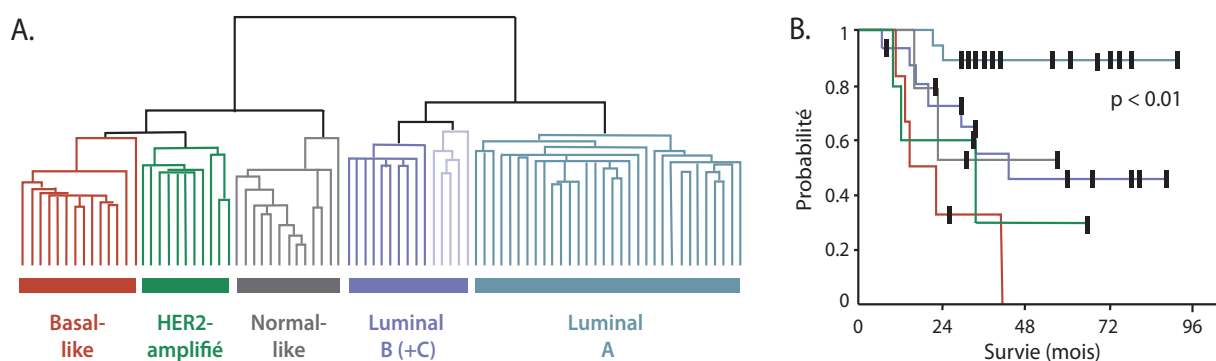


Figure 38 : Identification des sous-types moléculaires de cancers du sein

A. Le clustering hiérarchique d'une cohorte d'environ 70 tumeurs du sein, sur la base de l'expression de plus de 1000 gènes, a permis l'une des premières classifications moléculaires des tumeurs du sein en 5 groupes distincts. B. Cette classification moléculaire a avant tout une valeur pronostique : les différents sous-groupes de tumeurs présentent une survie globale à 5 ans significativement différente. Cette classification moléculaire permet en particulier de distinguer deux sous-groupes de tumeurs de type Luminal, d'évolution très différentes. (Adapté de Sørlie 2001 & Gruver 2011)

Une **réconciliation de l'approche immunohistochimique** conventionnelle a été proposée pour ces différentes signatures : on classe à présent les cancers du sein en trois groupes, les tumeurs lumineales (ER+/PR+), les tumeurs HER2-enrichies (HER2+) et les tumeurs *basal-like* (ER-/PR-/HER2-) (**FIGURE 39**).

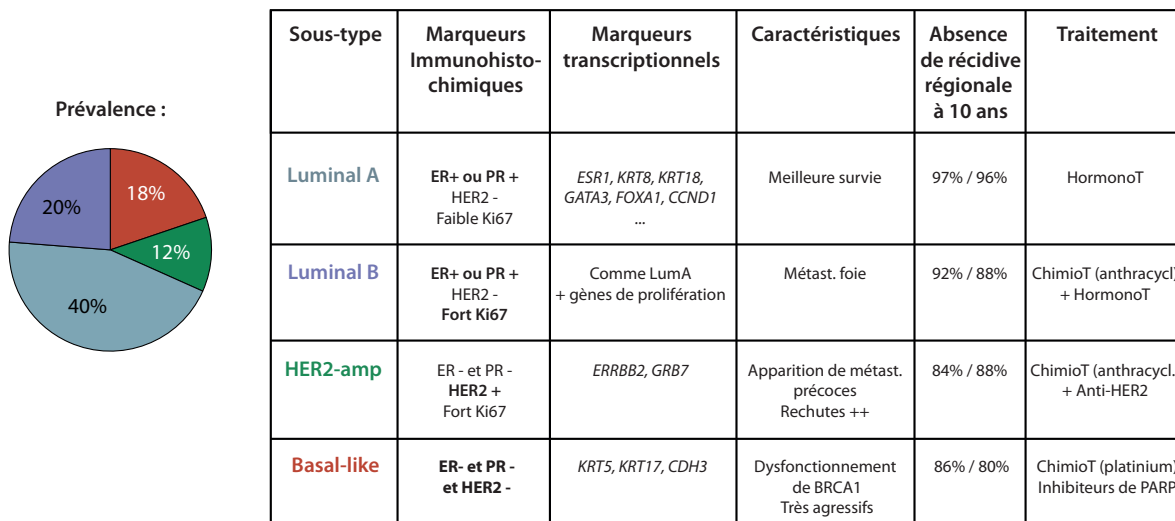


Figure 39 : Correspondance entre les différentes classifications

Ce tableau illustre les correspondances entre les sous-types moléculaires de cancers du sein, les marqueurs immunohistochimiques considérés en routine et les caractéristiques cliniques de ces tumeurs. L'espérance de survie sans récurrence régionale à 10 ans est donnée après traitement standard par une chirurgie conservatrice ou une mastectomie totale. (Selon Gruver et al. 2011, données de survie sans rechute d'après Voduc et al. 2010)

Notons cependant que si les appellations triple-négatif et *basal-like* sont souvent considérées comme synonymes, ces deux désignations ne concordent qu'imparfaitement. Pour illustrer ceci, j'ai croisé les données transcriptomiques et d'IHC concernant la cohorte de tumeurs du sein du TCGA : environ **20% des tumeurs triple-négatives ne sont pas basal-like et 30% des tumeurs basal-like ne sont pas triple-négatives.**

Avec l'évolution des techniques de séquençage, des classifications moléculaires de plus en plus pointues ont révélé de **nouvelles strates de la diversité des cancers du sein**, en particulier au sein du sous-groupe hétérogène des tumeurs basal-like ou triple négative. Par exemple, une étude suisse analysant des données d'expression par PCA et clustering hiérarchique a identifié le sous-type « **Apocrine** » de tumeurs du sein ER-/PR-, caractérisé par l'expression du récepteur aux androgènes et de marqueurs luminaux (Farmer et al. 2005). Dans la foulée, l'analyse intégrative de multiples données transcriptomique et le développement de modèles d'apprentissage (*machine-learning*) a permis l'identification d'une seconde entité de cancers ER-/PR-, caractérisée par une forte **signature interféron** et de réponse immunitaire et associée à un meilleur pronostic que les autres tumeurs *basal-like* (Hu et al. 2006, Teschendorff et al. 2007). Une autre sous-catégorie importante est celle du groupe « **Claudin-low** » identifié en 2007 par Herschkowitz et al. Ces tumeurs sont caractérisées par la répression de gènes impliqués dans l'adhésion cellulaire (*Claudin 3, 4 et 7, E-cadherin* notamment), par la surexpression de gènes impliqués dans l'EMT, et présentent des caractéristiques de cellules souches mammaires

(Herschkowitz et al. 2007). Ces tumeurs représentent jusqu'à 30% des cancers triple-négatifs et sont de moins bon pronostic que les tumeurs de type *basal-like* (Prat 2010). Enfin, certaines classifications ont utilisé des données épigénétiques : les travaux de Dedeurwaerder et al. rapportent l'existence de nouveaux sous-groupes de cancers du sein, identifiés grâce à l'analyse de données de méthylation de l'ADN. Ces sous-groupes n'étaient pas détectables dans les approches de classification reposant sur l'expression des gènes (Dedeurwaerder et al. 2011).

Enfin, nous terminerons en évoquant **les travaux de Lehmann et al. en 2011**, qui ont achevé de démontrer **l'éclatement du groupe des tumeurs triple-négative** (Lehmann et al. 2011). En analysant 21 jeux de données transcriptomiques publiques (microarrays), comprenant près de 600 tumeurs triple-négatives, les auteurs ont pu définir 6 sous-types moléculaires (**FIGURE 40A**). Le sous-type Basal-like 1 (BL1) : il se caractérise par une signature de gènes impliqués dans le cycle cellulaire et la réparation des dommages à l'ADN. Le sous-type Basal-like 2 (BL2) : proche des BL1, mais présentant un enrichissement de marqueurs myoépithéliaux et une signature de la réponse aux facteurs de croissance. Le sous-type Immunomodulatory (IM), qui se démarque par l'expression de cytokines et de gènes engagés dans la réponse immunitaire. Il correspondrait plus ou moins au groupe *interferon-rich* identifié par Hu et Teschendorff. Le sous-type Mesenchymal (M), qui exprime une signature EMT ainsi que des facteurs de croissance ; et le sous-type Mesenchymal stem-like (MSL), proche du sous-type M mais s'en distinguant par l'expression de gènes associés aux cellules souches et la faible expression de gènes impliqués dans la prolifération. Ce sous-groupe ressemblerait éventuellement à la définition du groupe des *claudin-low*. Enfin le sous-type Luminal androgen receptor (LAR), qui se caractérise par l'expression de marqueurs luminaux et par une signature typique de la signalisation du récepteur aux androgènes. Il pourrait correspondre, au niveau moléculaire, au sous-groupe apocrine. Les travaux de Mayer en 2014, recoupant la classification de Lehmann avec celle de Sorlie, a montré que le sous-type LAR est composé à 75% de tumeurs appartenant au groupe HER2-enrichies (Mayer et al. 2014). En 2015, d'autres approches ont légèrement modifié cette classification, en convergeant vers une classification en 4 ou 3 sous-types plus stables que ceux de la classification de Lehmann de 2011 (Burstein 2015, Jezequel 2015). L'année suivante, Lehmann révisé sa première classification en démontrant que certains sous-types traduisent en réalité l'hétérogénéité cellulaire et la composition du microenvironnement immunitaire, plutôt qu'un phénotype caractéristique des cellules tumorales elles-mêmes. Ainsi, le sous-type de Lehmann IM serait dû à une infiltration immunitaire importante, et le sous-type MSL traduirait la présence d'un stroma développé. (Lehmann et al. 2016). Les tumeurs appartenant aux sous-types IM et MSL seront donc reclassifiées selon une version ajustée des sous-types de Lehmann, ne comprenant plus que 4 sous-types moléculaires : BL1, BL2, M et LAR (**FIGURE 40B**).

La multiplicité des classifications proposées pour les cancers du sein triple-négatif traduit d'une part le caractère nébuleux de cette sous-catégorie de cancers, véritable boîte noire initialement définie par défaut, mais aussi d'autre part le vif intérêt de la communauté scientifique pour ces cancers encore mal définis. De fait, l'hétérogénéité à la fois clinique et biologique des cancers triple-négatifs représente certainement **un obstacle majeur pour l'identification de nouvelles cibles thérapeutiques**.

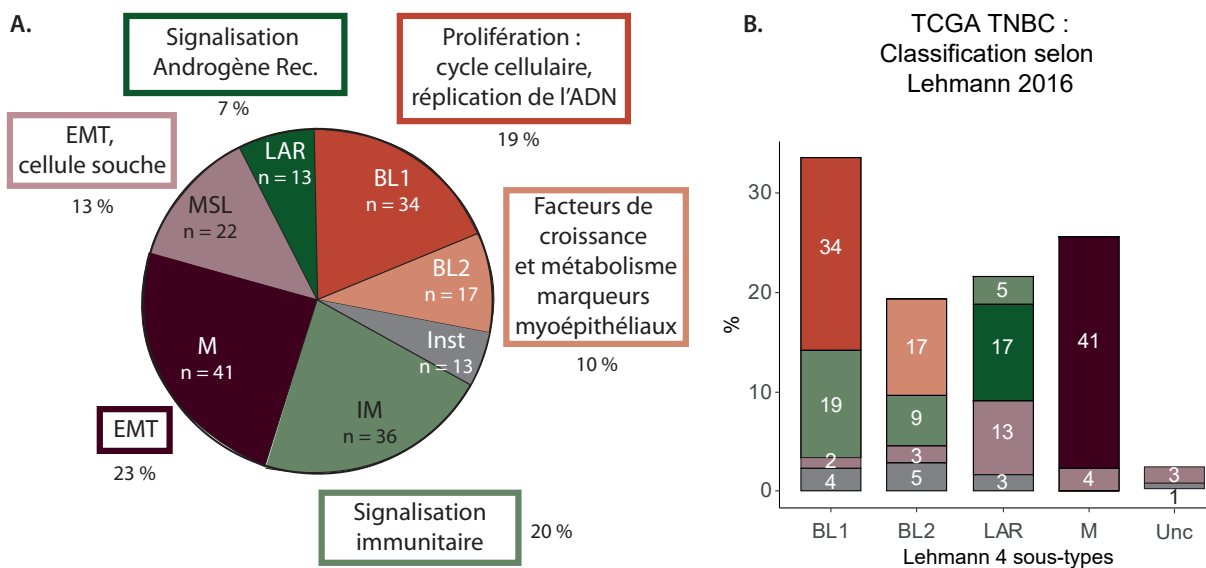


Figure 40 : Evolution des classifications moléculaires des tumeurs triple-négatives

A. La classification des 180 tumeurs du sein triple-négatives du projet TCGA, selon les critères définis par Lehmann et al. en 2011, nous permet d'identifier 6 sous-groupes moléculaires, dont les caractéristiques sont rappelées. Il existe un 7e groupe instable (Inst, 7 %), représenté en gris sur le graphique.

4. Classification des mêmes 180 tumeurs du sein triple-négatives du TCGA, cette fois selon les critères définis par Lehmann et al. en 2016. La couleur des rectangles correspond à la classification de 2011 (voir A) et leur taille est proportionnelle au nombre de tumeurs.

(Données du TCGA, analyse réalisée selon de Lehmann 2011, Lehmann 2016, Gerratana 2018)

Partie III

Quelles sont les propriétés des différents sous-types de cancers du sein ?

1. Origine cellulaire et moléculaire

Deux concepts peuvent expliquer cette hétérogénéité des tumeurs (FIGURE 41). Dans le premier modèle (**modèle génétique ou moléculaire**), la cellule d'origine des sous-types tumoraux est la même, mais différentes mutations entraînent différents phénotypes tumoraux. Dans le second modèle (**modèle cellulaire**), les différents sous-types tumoraux proviennent de la transformation de cellules d'origine distinctes dans la hiérarchie de différenciation du tissu sain (Visvader et al. 2011). Ces deux modèles ne sont pas exclusifs : au contraire, ils peuvent jouer tous deux à différents moments de l'oncogenèse pour déterminer le comportement de la tumeur.

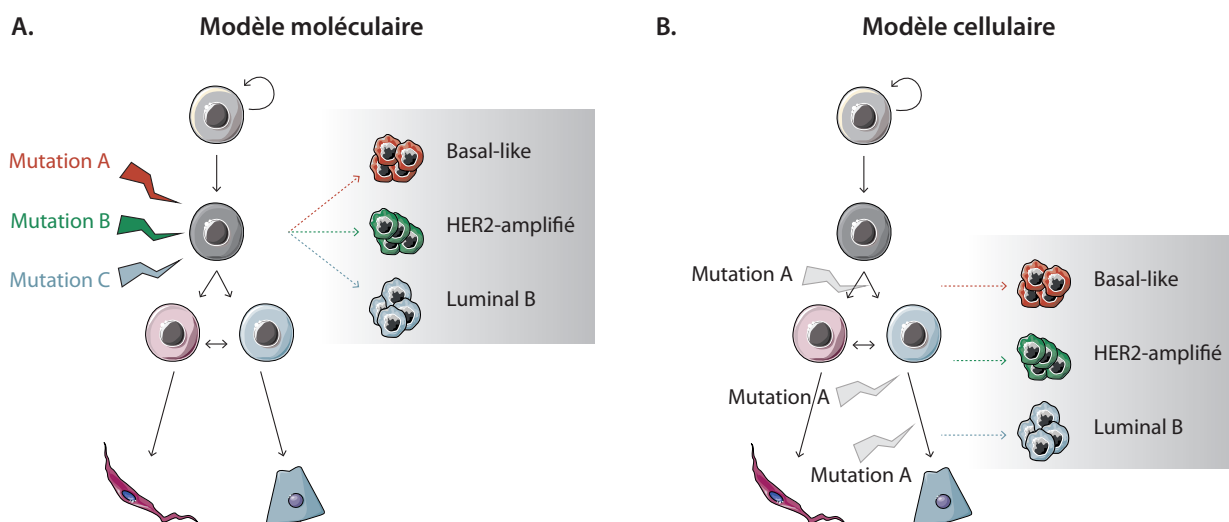


Figure 41 : Deux modèles pour expliquer la diversité des sous-types de cancers

A. Dans le modèle moléculaire, la nature des mutations génétiques (et épigénétiques) détermine principalement le phénotype de la tumeur. Des mutations différentes donneront des sous-types tumoraux différents.

B. Dans le modèle cellulaire, différents sous-types cellulaires au sein de la glande mammaire normale seront à l'origine des différents sous-types tumoraux. (Selon Visvader 2011)

Dans le cas des cancers du sein, l'hypothèse cellulaire est soutenue par plusieurs niveaux de preuves. Des éléments corrélatifs tout d'abord : les profils moléculaires des différents sous-types de tumeurs du sein et des différentes populations cellulaires de la glande mammaire se ressemblent et peuvent s'apparier. Pour ne citer qu'elle, la signature PAM50 évalue dans les tumeurs l'expression de transcrits qui sont des marqueurs des différentes cellules de la glande mammaire normale (ie. *FOXA1*, *PGR*, *ESR1*, *KRT14*, *KRT5*, *EGFR*, *FOXC1*, *MIA*). Ainsi, la signature des cellules souches mammaires adultes est très corrélée avec le profil d'expression des tumeurs *claudin-low*, tandis que la signature des progéniteurs luminaux est proche de celle des tumeurs du sous-type *basal-like* (Lim et al. 2009, Prat et al. 2010). Des éléments cliniques ensuite : les glandes mammaires saines de patientes mutées *BRCA1*, et qui développeraient généralement un cancer de type *basal-like*, présentent un enrichissement en progéniteurs luminaux (Lim et al. 2009). Des éléments expérimentaux enfin : les modèles murins mutés pour *Brca1* et *p53*, développent plutôt des tumeurs mammaires de sous-type *basal-like* ; et celles-ci émergent plus fréquemment de la transformation de progéniteurs luminaux plutôt que d'une autre cellule (Molyneux et al. 2010, Bai et al. 2012).

Les altérations génétiques contribue également à l'hétérogénéité des cancers du sein. Nombre de **variants somatiques sont communs aux différents sous-types** de cancers du sein (pour revue, voir par exemple [Ellis & Perou 2013](#)) : par exemple, les mutations perte de fonction de *P53* et les mutations activatrices de *PIK3CA* sont fréquemment retrouvées dans tous les sous-types de cancers du sein. Toutefois, certains patterns propres à chaque sous-type peuvent être dégagés, que nous allons décrire.

2. L'instabilité génomique est modérée (Luminal A) à forte (basal-like)

Les cellules cancéreuses modifient les grandes voies de réparation afin de maintenir un certain degré d'instabilité génomique, sans compromettre leur survie et en demeurant capable de supporter un niveau de stress génotoxique important. Dans les cancers du sein, le profil génomique des tumeurs mammaires est aussi hétérogène : grossièrement, on peut dire que **le degré d'instabilité génomique est négativement corrélé avec le pronostic des tumeurs** ([Vincent-Salomon et al. 2015](#)).

Les tumeurs de sous-type **Luminal A présentent un profil génétique relativement simple**, avec très peu de réarrangements chromosomiques. La principale altération retrouvée dans ce sous-type sont des gains affectant les chromosomes 1p et 16p, ainsi que la perte de 16q ([Ciriello et al. 2013](#)). En revanche, les tumeurs **Luminal B ont une instabilité génomique plus importante**, avec un profil d'altérations spécifiques appelé **amplifier** ([Kwei et al. 2010](#)). Ce profil est caractérisé par de nombreuses amplifications génétiques locales, regroupées sur un ou plusieurs bras chromosomiques. La plupart de ces amplifications impliquent des gènes de prolifération, tels que *FGFR1*, *MYC*, *CCND1*, *MDM2* et *ZNF214* ([Gatza 2014](#)) Les tumeurs **HER2-enrichies** ont un paysage génomique très semblable aux tumeurs Luminal B : on retrouve ce profil *amplifier*. Toutefois, et à la différence des tumeurs Luminal B, **l'amplification 17q12** (*ERBB2*) est un événement majeur de leur oncogénèse.

Enfin, les tumeurs **basal-like ont un profil génomique extrêmement complexe et remanié**, incluant de multiples pertes, gains et petites duplications en tandem, produisant un génome très fragmenté avec de nombreuses altérations du nombre de copies. Dans ce groupe de tumeurs, on retrouve fréquemment la **perte de fonction de BRCA1**, un facteur intervenant dans la réparation des cassures double-brin de l'ADN par recombinaison homologue ([Scully et al. 2014](#)), qui prédisposent au développement d'un cancer du sein de sous-type *basal-like* ([Lim et al. 2009](#), [Molyneux et al. 2010](#)). **Il existe de plus un phénotype spécifique appelé « BRCAness »**, que possèdent la **majorité des tumeurs basal-like** ([Turner et al. 2004](#), [Lips et al. 2013](#), [Lord et al. 2016](#)). Il décrit une situation où l'on observe un défaut de la voie de réparation de l'ADN par recombinaison homologue, mais sans mutation germinale de *BRCA1/2*.

3. Les réseaux de régulation transcriptionnels miment ceux de la cellule d'origine

Du côté de la glande mammaire saine, les cellules myoépithéliales, les progéniteurs luminaux et les cellules luminales présentent des signatures transcriptomiques spécifiques, sous le contrôle de voies de signalisation et de facteurs de transcriptions particuliers (**FIGURE 42**).

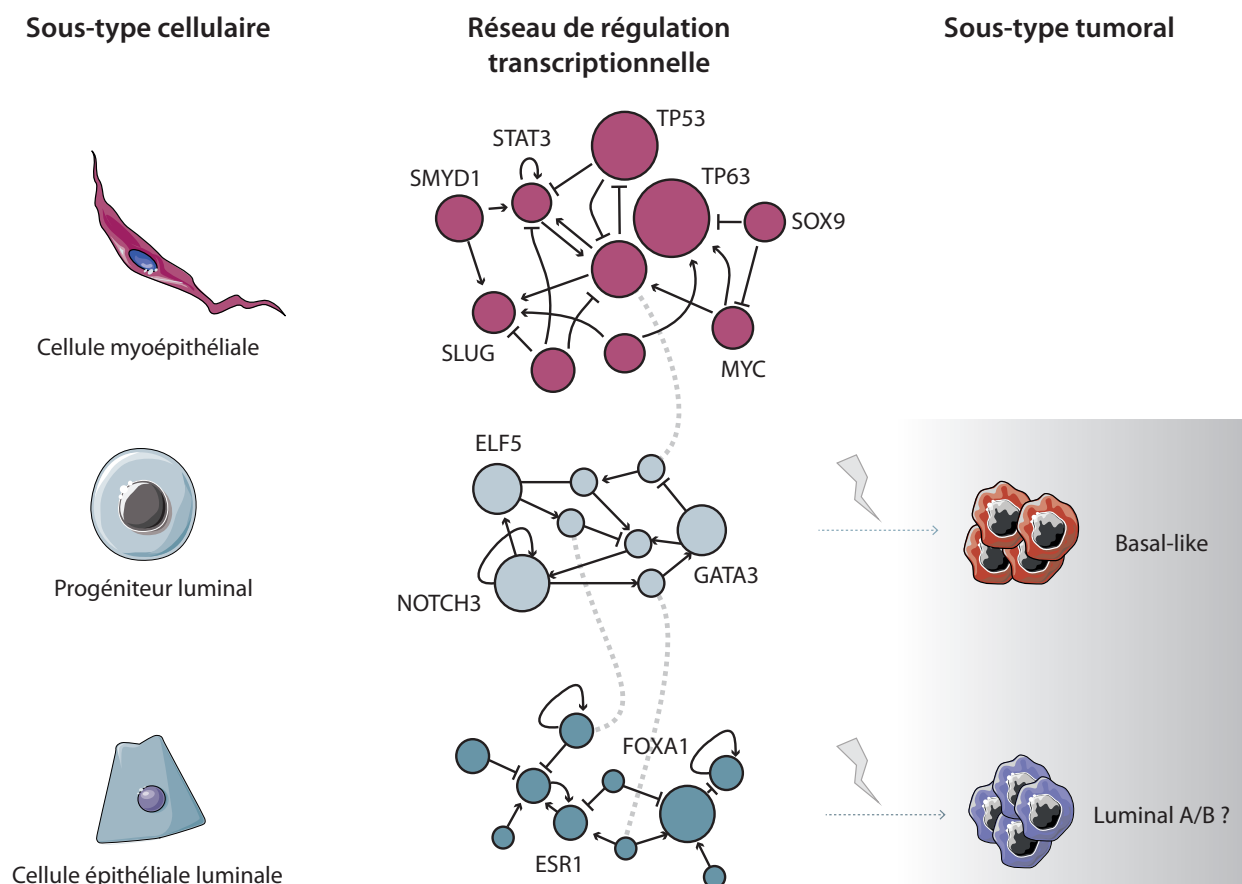


Figure 42 : Réseaux de régulation des sous-types cellulaires mammaires

Il existe des facteurs de transcription communs aux différentes populations cellulaires de la glande mammaire, représentés par une ligne pointillée grise. Les facteurs de transcription spécifiques de chaque sous-population ont été soulignés. Les réseaux de régulation entre les différents facteurs de transcription n'ont pas toujours été respectés, pour simplifier la représentation. (Adapté de Pellacani 2016)

Ces facteurs de transcription « clés » sont notamment : pour les cellules myoépithéliales, SLUG, **TP63**, TP53, SOX9, STAT3, **MYC** et TAZ; pour les cellules luminales : CEBPB et NOTCH3, critiques pour assurer l'engagement des progéniteurs dans la lignée luminales (Raouf et al. 2008), ainsi que **GATA3**, **ELF5** (Chakrabarti et al. 2012) pour les progéniteurs luminaux. Les facteurs **FOXA1** et **ESR1** semblent eux être indispensables à la différenciation des cellules luminales matures (Theodorou et al. 2013, Pellacani et al. 2016). Enfin pour les cellules souches mammaires : la voie **WNT** semble être importante pour assurer le maintien de ce compartiment (van Amerongen 2012).

D'autres voies importantes sont celles du TGF-beta (Kahata et al. 2017), ainsi que la voie Hippo (Pelissier et al. 2014, Britschgi et al. 2017). **Chacune des voies impliquées dans le développement normal de la glande mammaire peut être l'objet d'altérations dans les cancers du sein**, et ces altérations sont fréquemment associées au développement d'une tumeur présentant la même signature moléculaire que le sous-type cellulaire normal correspondant (Howars & Ashworth 2006, Pellacani et al. 2019). Par exemple, les altérations de FOXA1 (1-3% des cancers du sein) concernent principalement les tumeurs ER+ Luminal (TCGA 2012, Nik-Zainal et al. 2016), et l'activation aberrante d'ERα (70% des tumeurs du sein) est préférentiellement associée aux tumeurs lumineales (Magnani & Lupien 2014). Au contraire, l'amplification de l'oncogène *MYC* (plus de 15% des cancers du sein) est associée aux tumeurs ER-, plutôt de sous-type non-luminal (Nik-Zainal et al. 2016, Poli et al. 2018).

4. Des facteurs épigénétiques impliqués dans la différenciation et la plasticité des cancers du sein

Si les différents sous-types tumoraux proviennent de différentes cellules d'origines, les cellules cancéreuses hériteraient des épigénomes de ces différents sous-types cellulaires. Cette **stabilité des épigénomes** est particulièrement criante lorsque l'on s'intéresse aux travaux de classification des tumeurs : les classifications épigénétiques, en particulier basées sur la méthylation de l'ADN, sont celles qui montrent la plus grande robustesse et conservation entre le tissu (ou la cellule) d'origine et la tumeur (Hoadley et al. 2018), et ont permis de disséquer finement certains sous-groupes de tumeurs du sein. Cependant, la transformation tumorale s'accompagne également de modifications importantes des paysages épigénétiques, et les cancers du sein n'échappent pas à cette règle (Locke & Clark 2012).

Comme dans le cas général, les cellules cancéreuses du sein présentent une **hypométhylation globale**, qui participe à l'instabilité génomique. (Robertson et al. 2005, Jones & Baylin 2002) et à l'activation d'oncogène : par exemple, l'hypométhylation de *CDH3* conduit à la surexpression de la cadhérine P et corrèle avec le caractère invasif des tumeurs (Paredes et al. 2005). Parallèlement, l'**hyperméthylation locale de certains promoteurs riches en CpG de gènes** réprime l'expression de gènes suppresseurs de tumeurs (Jones, Issa & Baylin 2016 ; pour revue voir Hinshelwood & Clark 2008). Nous avons déjà abordé l'hyperméthylation d'un des allèles du gène *BRCA1*, qui associé à la mutation du second allèle conduit à la perte de fonction totale de la protéine BRCA1 (Esteller et al. 2000). On peut également retenir la méthylation du promoteur de *CDH1*, identifiée comme l'un des mécanismes permettant la perte d'expression de l'E-cadhérine et la progression métastatique (Graff et al. 2000). Certaines stratégies thérapeutiques exploitent cette caractéristique : chez la souris, il a été démontré que le traitement par un inhibiteur de DNMT1 (5-AzaC) permet de diminuer la croissance des tumeurs mammaires et d'améliorer la survie globale (Pathania et al. 2016).

Les cellules cancéreuses des tumeurs du sein **détournent également les fonctions normales des régulateurs épigénétiques (FIGURE 43)**.

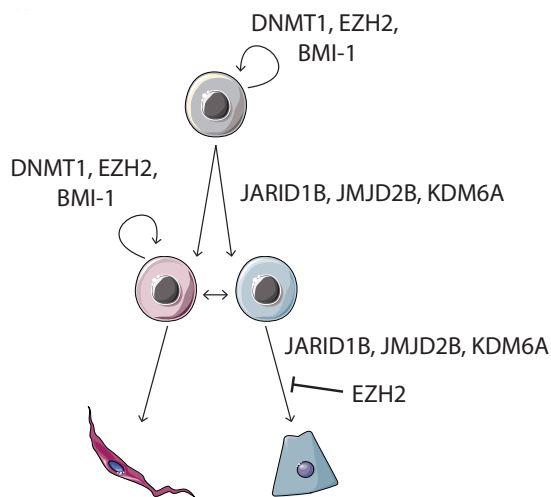


Figure 43 : Facteurs épigénétiques impliqués dans la différenciation de la glande mammaire

Chaque stade de différenciation des cellules de la glande mammaire nécessite l'action d'enzymes épigénétiques spécifiques, afin de sceller le destin cellulaire de chaque lignée. (Adapté de Holliday 2018 et de Locke & Clark 2012)

Par exemple, dans les cellules non-cancéreuses, EZH2 est nécessaire à la prolifération des progéniteurs luminaux (Pal et al. 2013). Or, tout comme les progéniteurs luminaux, les cancers du sein *basal-like* de haut grade surexpriment EZH2, ce qui soutient leur prolifération (Kleer et al. 2003). De même, la protéine du complexe PRC1 BMI1, nécessaire aux propriétés d'auto-renouvellement des cellules souches mammaires dans la glande normale, est surexprimée par certaines tumeurs basal-like particulièrement agressives où elle stimule l'EMT et la prolifération (Guo et al. 2010, Paranjape et al. 2014).

Du côté des cancers luminaux, l'histone déméthylase JARID1B promeut la différenciation des progéniteurs luminaux (Zou et al. 2014), et est fréquemment amplifiée et surexprimée dans les cancers du sein de sous-type luminal (Yamamoto et al. 2014). Ces exemples soulignent l'importance du jeu équilibré des enzymes de remodelage de la chromatine : dans le contexte cellulaire de la tumeur, ces facteurs épigénétiques agissent comme des oncogènes, là où leurs fonctions participent au développement harmonieux de la glande mammaire dans les cellules non-cancéreuses.

Les anomalies épigénétiques des cellules cancéreuses offrent également des possibilités thérapeutiques. Par exemple, l'hyperméthylation du promoteur de *ESR1* entraînant la perte d'expression du récepteur ER alpha est observée chez environ 20% des patientes progressant sous hormonothérapie. L'utilisation d'inhibiteurs des HDACs permettrait de restaurer l'expression d'ER alpha et de lever la résistance ; cette approche a démontré son succès *in vitro* (Shiino et al. 2016) mais les essais cliniques engagés en ce sens se sont révélés infructueux, suggérant l'existence de mécanismes épigénétiques supplémentaires assurant la répression de *ESR1*. Une autre approche s'intéresse aux inhibiteurs de BET, qui ciblent les protéines reconnaissant l'acétylation des histones. Ces traitements montrent par exemple une bonne efficacité *in vitro*, particulièrement dans les cancers du sein *basal-like*. Ils semblent agir en réprimant l'expression des gènes spécifiques de la lignée basale, en inhibant notamment l'interaction entre BRD4 et FOXO1 (Nagarajan et al. 2016). De façon intéressante, ces inhibiteurs de BET semblent stimuler la différenciation des cellules cancéreuses *basal-like* vers un phénotype plus luminal. Enfin; les thérapies épigénétiques peuvent sensibiliser les cellules cancéreuses au système immunitaire, et potentialiser ainsi les effets des inhibiteurs des points de contrôle immunitaires (Dunn et al. 2017). Ces traitements seraient particulièrement adaptés aux cancers du sein triple-négatifs, où les inhibiteurs de checkpoint ont une efficacité très limitée par rapport à d'autres tumeurs solides (Emens et al. 2018).

5. Une hétérogénéité intratumorale variable en fonction du sous-type

Similairement au schéma général des cancers, l'hétérogénéité inter-tumorale des cancers du sein se double d'une hétérogénéité intrinsèque de la tumeur.

D'une part, la composition du **microenvironnement tumoral** influence le comportement des cellules cancéreuses de la glande mammaire. Les travaux précurseurs de DeCossé et al. ont démontré dès les années 70 la capacité des cellules du stroma à réguler la croissance et les différenciation des cellules cancéreuses du sein (DeCossé et al. 1973, DeCossé et al. 1975) ; depuis de nombreuses études ont solidement établi que des caractéristiques tumorales sont influencées par les cellules du microenvironnement tumoral. Réciproquement, l'équipe de Kornelia Polyak a analysé en détail les paramètres moléculaires de chaque type cellulaire composant le tissu tumoral ou la glande mammaire normale, et a pu montrer l'importance des changements transcriptomiques survenant dans les cellules du microenvironnement. En particulier, les cellules myoépithéliales présentes dans le tissu tumoral surexpriment les chemokines CXCL14 et CXCL12, qui stimulent par communication paracrine la prolifération, la migration et l'invasion des cellules tumorales (Allinen et al. 2004).

D'autre part, les cancers du sein présentent une **hétérogénéité tumorale intrinsèque (FIGURE 44-45)**, pour lequel ont été démontré à la fois la composante clonale et la composante cellule souche cancéreuse (CSC).

L'hétérogénéité clonale des tumeurs du sein est révélée notamment par la survenue de résistances thérapeutiques. Par exemple, l'analyse par single-cell DNA-seq de 20 tumeurs du sein triple-négative montre l'existence, avant tout traitement, de clones tumoraux génétiquement distincts, responsables de la résistance aux traitements néoadjuvants : les traitements anti-cancéreux viennent en réalité sélectionner et amplifier les clones tumoraux résistants (Kim et al. 2018). Même résultat dans des lignées cellulaires de cancer du sein ER+ : des sous-clones génétiquement spécifiques, résistants aux traitement anti-ostrogéniques (fulvestrant et tamoxifène), préexistent de façon minoritaires et se trouvent sélectionnés par le traitement (Hinohara et al. 2018). La considération de l'hétérogénéité clonale des tumeurs du sein peut donc avoir d'importantes conséquences sur leur prise en charge et l'utilisation des combinaisons thérapeutiques appropriées. Plus surprenant encore, l'analyse transcriptomique par nanogrid à l'échelle de noyaux cellulaires isolés d'une tumeur ER-/PR-/HER2- a conclu que, si la majorité des cellules présentent une signature moléculaire typique des tumeurs *basal-like*, un pourcentage significatif de noyaux cellulaires tumoraux sont catégorisés comme HER2-enrichies, luminal A, B, ou *normal-like*, indiquant la présence conjointe de différents phénotypes cellulaires mimant les différents sous-types tumoraux au sein d'une même tumeur (Gao et al. 2017). De plus amples analyses sont nécessaires pour concilier le modèle cellulaire de développement des sous-types de cancers du sein avec un tel degré d'hétérogénéité intratumoral. L'hétérogénéité génétique de la tumeur peut également participer aux processus métastatiques, en permettant l'émergence de cellules cancéreuses aux propriétés plus mésenchymateuses. Grâce à des techniques de séquençage single-cell sur différentes régions d'une même tumeur, plusieurs études ont montré que les clones

généétiques responsables de l'invasion tumorale peuvent apparaître très tôt au cours de l'histoire évolutive des tumeurs du sein, dès le stade carcinome canalaire *in situ* (Martelotto et al. 2017, Casasent et al. 2018). Toutefois, étant donné que la majorité des carcinomes canauxaux *in situ* n'évolueront pas vers une maladie métastatique, il est clair que la présence de ces clones plus invasifs ne suffit pas à déterminer la progression tumorale : d'autres facteurs, et en particulier le rôle du microenvironnement et de la réponse immunitaire, contrôlent également ce processus (Gil Del Alcazar et al. 2017, Dongre & Weinberg 2018)

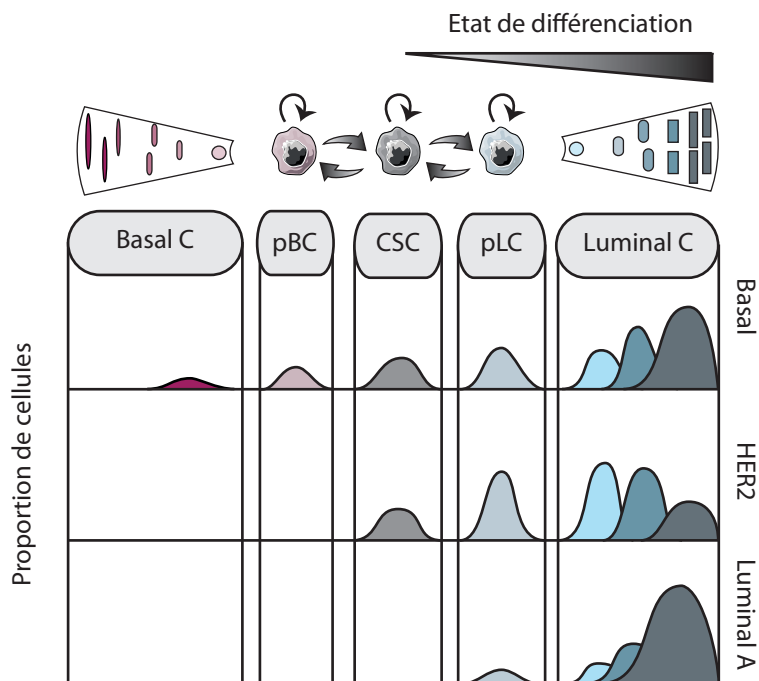


Figure 44 : Variation des proportions de cellules souches cancéreuses et de leurs descendances différenciées selon le sous-type de tumeurs du sein

Dans ce modèle, les différents sous-types moléculaires des cancers du sein sont caractérisés par différentes proportions de cellules souches cancéreuses (CSC) engagées vers un état mésenchymateux (ou basal, pBC) versus épithélial (ou luminal, pLC), ainsi que par des proportions variées de descendants plus ou moins différenciés (Basal C vs. Luminal C) selon la hiérarchie cellulaire observée dans le développement de la glande mammaire saine. (Adapté de Brooks 2015)

L'existence d'une (voire plusieurs) population CSCs dans les tumeurs du sein est maintenant bien établie. Les CSCs mammaires ont été identifiées sur la base de l'expression de marqueurs de surface (CD44+/CD24-) (Carrasco et al. 2015, Mansoori et al. 2017), et la forte expression de l'aldéhyde déshydrogénase, une enzyme contrôlant la différenciation cellulaire (Ginestier et al. 2017). Les CSCs du sein proviendraient de la transformation directe des cellules souches adultes normales de la glande mammaire. Elles sont capable d'effectuer une transition entre un état épithélial plus prolifératif et un état mésenchymateux plus quiescent et invasif, générant leurs descendances cellulaires respectives. Les proportions des différentes populations de CSCs et de leurs descendances différenciées varient selon les sous-types moléculaires : le sous-type luminal A contient principalement des cellules tumorales épithéliales bien différenciées ; au contraire, le sous-type *basal-like* contient des cellules épithéliales mais également des cellules tumorales mésenchymateuses et des cellules exprimant des marqueurs souches.

L'épigénétique participe donc aussi à l'hétérogénéité intratumorale des cancers du sein. Des études récentes ont décrit l'évolution de la méthylation de l'ADN et des marques d'histones au cours de la progression de la maladie. Par exemple, l'acquisition d'un phénotype de résistance aux traitements par chimiothérapie des tumeurs du sein triple-négatif s'accompagne de l'émergence de sous-populations tumorales, polyclonales d'un point de vue génétique mais présentant une signature épigénétique commune. Cette signature se distingue par la redistribution de la marque épigénétique répressive H3K27me3, à l'origine de l'activation d'un programme transcriptionnel spécifique aux cellules tumorales persistant sous chimiothérapie. Des modifications épigénétiques sont donc suffisantes pour réguler l'activation différentielle de voies impliquées dans la résistance à la chimiothérapie dans les cancers du sein triple-négatif (telles que les voies Hedgehog, WNT, TGF- β) (Marsolier et al. 2021). Ces études soulignent l'importance d'une cartographie précise des altérations épigénétiques dans les tumeurs du sein, permettant d'identifier des marqueurs pronostiques de réponse au traitement et éventuellement de nouveaux leviers thérapeutiques pour ces cancers.

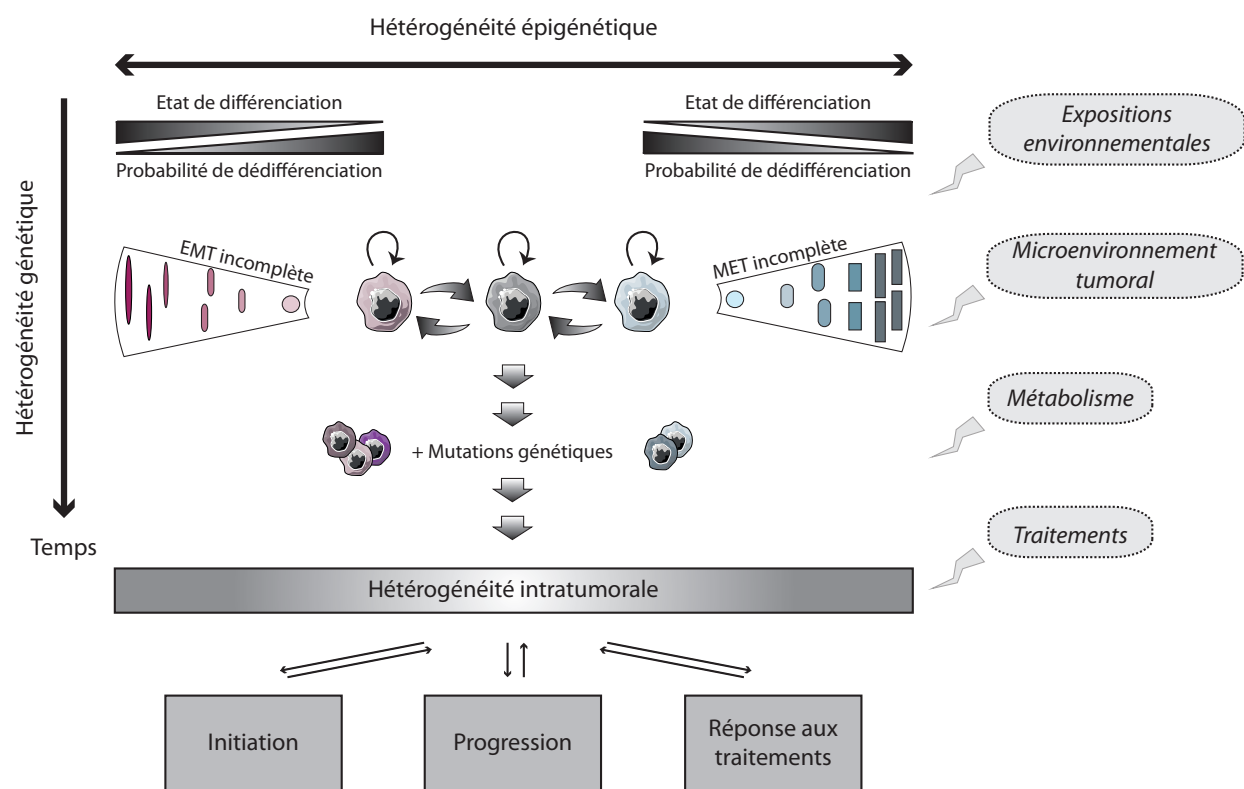


Figure 45 : Hétérogénéité génétique et épigénétique dans les cancers du sein

Différentes cellules souches cancéreuses existent au sein d'une tumeur mammaire : l'une est de phénotype mésenchymal et quiescente (en rose, pBC, CD44+/CD24-), l'autre est épithéliale et proliférative (en bleu, pLC, ALDH+). Une population de cellules souches possédant ces deux caractéristiques a été isolée (en gris), mais il n'est pas encore clair s'il s'agit d'une population réelle ou d'un état plastique transitoire. Dans ce modèle, on peut également retrouver des cellules présentant des caractéristiques incomplètes d'EMT ou de MET, selon leur niveau de différenciation. De plus, l'accumulation de mutations génétiques au cours de l'évolution de la tumeur sera à l'origine de nouveaux clones de cellules souches cancéreuses et de leurs descendants. Ces processus génétiques et épigénétiques d'évolution polyclonale des tumeurs sont influencés par de multiples facteurs, qui agissent sur les phénotypes tumoraux à différentes échelles de temps. Tous ces phénomènes peuvent se combiner en amplifiant encore l'hétérogénéité intratumorale, qui détermine in fine les mécanismes biologiques d'initiation, de progression et de réponse aux traitements. (Adapté de Brooks 2015, Hinohara & Polyak 2019)

6. Les gènes Cancer/Testis dans les cancers du sein

Les gènes C/T ont été notre porte d'entrée pour étudier ces modifications épigénétiques des tumeurs du sein. L'activation des gènes C/T dans ces cancers a été étudiée en particulier pour son potentiel d'application pronostique et thérapeutique, étant donnée la spécificité d'expression des gènes C/T. Sur la base de ces études, plusieurs résultats sont à souligner.

Les tumeurs du sein ont une fréquence d'expression des gènes C/T intermédiaire (Scanlan et al. 2004), auquel s'ajoute des variations selon le sous-type tumoral. Plusieurs groupes ont montré une **association entre le sous-type de cancer du sein et l'activation de certains gènes C/T**. En 2011, Curigliano et al. ont montré par IHC une expression significativement plus fréquente des gènes C/T MAGEA et NY-ESO-1 dans les tumeurs triple négative (26% et 18% respectivement) que dans les tumeurs ER-positives (10% et 4%) (Curigliano et al. 2011). Cette tendance des tumeurs triple-négatives à activer un plus grand nombre de gènes C/T que les tumeurs ER-positives se retrouve également dans une étude indépendante (Chen et al. 2011). Il serait donc utile de nuancer ce statut intermédiaire des tumeurs du sein dans l'activation des gènes C/T, et de le différencier selon le sous-type tumoral.

L'activation de certains gènes C/T a également une **valeur pronostic**. L'expression des MAGEs corrèle avec le stade TNM, la présence de métastases dans les ganglions lymphatiques juxta-tumoraux, la taille de la tumeur ainsi qu'une moins bonne survie globale (Lian et al. 2012, Ayyoub et al. 2014, Abd-Elsalam 2014). Cependant **l'expression préférentielle de ces gènes par les tumeurs triple-négatives n'a pas été prise en compte**, et peut être dans ce cas un facteur confondant important.

En ce qui concerne les fonctions des gènes C/T dans les tumeurs du sein, une étude de la littérature montre la diversité de celles-ci : **on retrouve toutes les grandes fonctions des gènes C/T dans la progression tumorale** exposées en II-2-4. Par exemple, l'activation de MAGEC2 induit une EMT dans des cellules cancéreuses de sein et favorise la dissémination métastatique des carcinomes mammaires (Yang et al. 2014). Certains gènes C/T, tels SPAG9, stimulent la prolifération des cellules tumorales du sein (Sinha et al. 2013) ; d'autres font des cibles intéressantes pour le développement d'immunothérapies, comme CT83 dans les tumeurs du sein triple-négatives (Paret et al. 2015).

Malgré ces résultats encourageants, peu d'études extensives des associations entre activation des gènes C/T et caractéristiques des tumeurs du sein ont été conduites. Lorsque cela a été fait, ces analyses sont restées principalement au niveau de l'analyse rétrospectives de données bioinformatiques et cliniques (voir par exemple : Grigoriadis et al. 2009 ; Holm et al. 2016), sans aller jusqu'à associer des études expérimentales permettant d'adresser la question du rôle fonctionnel de ces activations illégitimes. En revanche, certaines études exploratoires ont révélé des associations significatives entre sous-type tumoraux et expression différentielle d'oncogènes, et ont ainsi découvert le rôle joué par certains gènes C/T dans la progression tumorale, soutenu par des modèles expérimentaux convaincants (Watkins et al. 2015). Ces études ont renforcé notre conviction de l'intérêt d'une étude alliant prédictions bioinformatiques et modèles fonctionnels de l'activation des gènes C/T dans les cancers du sein.

Objectifs de la thèse

Mon travail de thèse s'articule autour de la question centrale de la régulation épigénétique des gènes C/T, des mécanismes à l'origine du maintien de leur répression dans les cellules somatiques et des événements conduisant à leur activation aberrante dans certaines cellules tumorales. A travers cette problématique s'entrecroise différents enjeux que nous avons évoqué au cours de cette introduction : la question du maintien des identités cellulaires et de leur altération par les processus oncogénétiques ; les relations entre signalisation cellulaire et contrôle épigénétique des gènes ; le rôle des modifications épigénétiques dans la transformation et la progression tumorale.

Plus précisément, les questions qui m'ont intéressées sont les suivantes :

- Peut-on identifier, par une approche exploratoire non biaisée, des gènes C/T spécifiquement activés dans certains contextes cellulaires ?
 - Suite à la modification expérimentale de l'activité d'acteurs signalétiques et épigénétiques ?
 - Par corrélation avec les contextes tumoraux dans le cas des cancers du sein ?
- Quels sont les voies de signalisation qui contrôlent la répression épigénétique des gènes C/T dans les cellules saines, et leur activation dans les cellules tumorales ?
 - Ces voies de signalisation sont-elles conservées entre cellules saines et cellules transformées ? Entre différents types tumoraux ?
 - Quel est le rôle de la conformation chromatinienne de la cellule d'origine dans l'activabilité des gènes C/T lors de la transformation ? Et celui des mutations oncogéniques particulières ?
- L'activation atypique de ces gènes dans les cellules tumorales est-elle uniquement un marqueur des anomalies épigénétiques survenant au cours du développement tumoral, ou les produits d'expression des gènes C/T jouent-ils un rôle dans la biologie des tumeurs ?
 - Etant donné que les partenaires habituels de ces protéines ne sont pas systématiquement co-exprimés dans les cellules cancéreuses, le rôle éventuel de l'activation des gènes C/T dans les tumeurs est-il différent de celui qui leur incombe dans les cellules germinales ? Est-il conservé entre les différents types tumoraux ?
 - Quel est la place de ces activations illégitimes dans la chronologie des événements transformants ?

J'ai voulu développer ces questions à travers deux projets complémentaires : le premier repose sur deux cribles expérimentaux dans un modèle cellulaire non cancéreux, pour aboutir sur la construction de modèles bioinformatiques prédictifs du comportement tumoral en extrapolant les résultats obtenus sur le modèle sain. Le second emprunte le chemin inverse, en partant d'un travail bioinformatique de

caractérisation et de prédiction de l'expression des gènes C/T dans les tumeurs du sein, résultats qui serviront de support à l'élaboration de modèles expérimentaux impliquant des lignées cellulaires précancéreuses de cellules épithéliales mammaires.

L'objectif fondamental de ma thèse est donc de découvrir de nouveaux mécanismes régissant l'expression ou la répression des gènes C/T, à la fois dans les cellules somatiques saines et dans les cellules tumorales, et d'interroger le rôle éventuel des produits d'expression de ces gènes dans les cancers. J'ai combiné des approches expérimentales et bioinformatiques, afin de tirer profit à la fois de l'expertise en biologie moléculaires et cellulaires du laboratoire mais également de l'immense quantité de données transcriptomiques et épigénétiques sur les tumeurs disponibles publiquement.

RESULTATS

Partie I.

Régulation de l'expression des gènes C/T dans les cellules non transformées

1. Définition du cadre expérimental

Dans la première partie de mon travail, j'ai rejoint un projet développé au laboratoire par une doctorante plus avancée, Ikrame Naciri. **Nous nous intéressons aux mécanismes verrouillant l'expression des gènes C/T dans les cellules saines**, et cherchions à identifier les voies de signalisations et les acteurs moléculaires responsables de leur répression. Pour ce faire, deux cribles génétiques ont été développés.

Le premier crible, s'intéressant aux voies de signalisation, fonctionnait sur le modèle du gain-de-fonction et exploitait une banque de 192 kinases impliquées dans la signalisation cellulaire, constitutivement activées. Pour ce faire, les kinases ont été fusionnées à une séquence de myristoylation, entraînant leur recrutement à la membrane plasmique où elles seront séquestrées et perpétuellement activées. Certaines de ces kinases sont connues pour être des oncogènes, telles que RET ou FGFR1. Cette banque a été utilisée dans d'autres études (Boehm 2007), et est constituée de vecteurs rétroviraux encapsulant les séquences des kinases recombinées.

Le deuxième crible concerne les modulateurs épigénétiques responsables de la répression des gènes C/T, et fonctionne cette fois sur le modèle de perte de fonction. Nous avons établi une banque de siRNA dirigés contre 160 modulateurs épigénétiques. Ces modulateurs sont des répresseurs transcriptionnels, tels que des lysines méthylases, des lysines désacétylases et les ADN méthyltransférases DNMT1, 3a, 3b, mais aussi des activateurs de la transcription qui pourraient jouer un rôle indirect dans la répression des gènes C/T.

Comme modèle cellulaire pour réaliser ces deux cribles, nous avons choisi **une lignée humaine de cellules non transformées** : il s'agit de la lignée de fibroblastes embryonnaires humain de poumon IMR90. Cette lignée présente l'avantage d'être bien caractérisée (de nombreuses données transcriptomiques et de méthylation sont disponibles publiquement), et d'être assez stable en culture. Cependant ce modèle présente certains inconvénients, et nous discuterons ces aspects litigieux : tout d'abord il s'agit d'une lignée embryonnaire et non adulte, dont la biologie obéit à des mécanismes particuliers. Ensuite il s'agit d'une lignée de fibroblastes, or la plupart des tumeurs auxquelles je me suis intéressé sont des carcinomes dérivés de la transformation de cellules épithéliales.

Le projet expérimental était le suivant : nous voulions soit infecter les IMR90 par les kinases activées, soit les transfecter avec les siRNA visant les modulateurs épigénétiques, en modulant l'activité d'un acteur à la fois. Les cellules ont été récoltées 4 jours après l'infection ou la transfection, l'ARN en a été extrait. Nous souhaitons mesurer l'effet de ces cibles sur l'expression d'un certain nombre de gènes C/T par Nanostring. En comptant les 192 kinases et les 160 siRNA, ainsi les conditions témoins non traitées, différents contrôles positifs (notamment une lignée dérivée des IMR90 et immortalisée par la télomérase, les SW39), des gènes de référence pour la normalisation des résultats et des sondes permettant de valider la diminution d'expression des modulateurs visés par siRNA, nous avons calculé que **nous pouvons analyser conjointement l'expression de 42 gènes C/T**.

Pour sélectionner les 42 gènes C/T dont nous allons analyser l'expression, nous nous sommes intéressés à plusieurs critères :

- **Nous voulions sélectionner des gènes C/T susceptibles d'être activés dans les cancers.** Pour se faire, l'équipe a utilisé la bibliographie disponible, en se basant principalement sur les travaux de l'équipe de Saadi Kochbin sur l'expression des gènes C/T dans les tumeurs du poumon.
- Nous nous intéressons particulièrement au **rôle de la méthylation** dans le contrôle de l'expression des gènes C/T : ont donc été privilégiés des gènes C/T présentant un îlot CpG dans leur promoteur et / ou une sensibilité aux traitements déméthylants avérée dans la littérature.
- Enfin, dans l'objectif d'identifier les mécanismes régulation de gènes C/T susceptibles d'être directement impliqués dans l'oncogenèse, et donc d'être des cibles thérapeutiques potentielles, nous avons utilisé les données bibliographiques disponibles pour sélectionner des gènes C/T présentant des **propriétés oncogéniques**.

2. Conclusions majeures du projet

Toute l'analyse primaire des résultats de Nanostring a été réalisée préalablement à ma thèse par le bioinformaticien de l'équipe, Olivier Kirsh. J'ai pu, sur la fin du projet, contribuer à l'analyse secondaire de ces résultats et à la réflexion autour de la meilleure façon de les présenter.

La conclusion principale de cette analyse est un résultat négatif intéressant : **pour la grande majorité des gènes C/T étudiés, ni l'activation d'une unique kinase signalétique, ni l'inhibition d'un seul modulateur épigénétique, ne suffit à surexprimer ces gènes C/T dans le modèle cellulaire étudié (ARTICLE : FIGURE 1 ET S1).**

Cependant, **un petit nombre de gènes C/T montrent un comportement différent, et semblent plus directement modulables par nos cribles** : l'exemple le plus flagrant est celui d'*ADAM12*, significativement induit par l'activation de la MAP kinase TAK1 (située dans la voie non-canonique du TGF beta) et par l'inhibition de l'histone acétyltransférase KAT2A (**ARTICLE : FIGURES 1-6 ET S1-S6**). Nous nous sommes donc concentrés sur ce gène C/T, et avons pu décortiquer les mécanismes moléculaires à l'œuvre dans les cellules non transformées, dans des lignées cellulaires de cancers, et dans des échantillons de tumeurs humaines. Ces résultats ont fait l'objet d'un article dont je suis co-première autrice, publié en 2019 dans le journal *Nucleic Acids Research*.

Genetic screens reveal mechanisms for the transcriptional regulation of tissue-specific genes in normal cells and tumors

Ikrane Naciri^{1,†}, Marthe Laisné^{1,†}, Laure Ferry¹, Morgane Bourmaud², Nikhil Gupta¹, Selene Di Carlo³, Anda Huna⁴, Nadine Martin⁴, Lucie Peduto³, David Bernard⁴, Olivier Kirsh^{1,*} and Pierre-Antoine Defossez^{1,*}

¹Univ. Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, 75013 Paris, France, ²INSERM U1132 and USPC Paris-Diderot, Hôpital Lariboisière, Paris, France, ³Unité Stroma, Inflammation & Tissue Repair, Institut Pasteur, 75724 Paris, France; INSERM U1224, 75724 Paris, France and ⁴Centre de Recherche en Cancérologie de Lyon, Inserm U1052, CNRS UMR 5286, Université de Lyon, Centre Léon Bérard, 69008 Lyon, France

Received January 08, 2019; Revised January 28, 2019; Editorial Decision January 29, 2019; Accepted January 30, 2019

ABSTRACT

The proper tissue-specific regulation of gene expression is essential for development and homeostasis in metazoans. However, the illegitimate expression of normally tissue-restricted genes—like testis- or placenta-specific genes—is frequently observed in tumors; this promotes transformation, but also allows immunotherapy. Two important questions are: how is the expression of these genes controlled in healthy cells? And how is this altered in cancer? To address these questions, we used an unbiased approach to test the ability of 350 distinct genetic or epigenetic perturbations to induce the illegitimate expression of over 40 tissue-restricted genes in primary human cells. We find that almost all of these genes are remarkably resistant to reactivation by a single alteration in signaling pathways or chromatin regulation. However, a few genes differ and are more readily activated; one is the placenta-expressed gene ADAM12, which promotes invasion. Using cellular systems, an animal model, and bioinformatics, we find that a non-canonical but druggable TGF- β /KAT2A/TAK1 axis controls ADAM12 induction in normal and cancer cells. More broadly, our data show that illegitimate gene expression in cancer is an heterogeneous phenomenon, with a few genes activatable by simple events, and most genes likely requiring a combination of events to become reactivated.

INTRODUCTION

The human body contains ~200 cell types, each characterized by a specific gene expression pattern. This pattern itself is determined by transcription factors, acting on a chromatin template rendered more or less permissive to their action by chromatin-modifying factors, such as DNA methyltransferases and demethylases, histone modifying enzymes, and nucleosome remodelers (1,2). These gene expression events are also influenced by cellular signaling pathways, which transmit the intracellular and extracellular signals that the cell is subjected to during development and during its normal life (3,4). A well-known example of extracellular signal is the cytokine Transforming Growth Factor β (TGF- β), which plays complex roles during development, immunity and cancer (5). Transcriptional regulation by chromatin-templated processes and cellular signaling have each been studied extensively individually, yet the interplay between these two processes has been harder to decipher. A few examples of kinase signaling cascades influencing chromatin status have been reported (6,7), but these findings have not been generalized.

Cancer cells show abnormalities in signaling and in chromatin regulation, leading to illegitimate gene expression, i.e. the expression of a gene in a tissue type where it is normally silenced (8). This illegitimate expression can contribute to tumorigenesis (9), however the inappropriate expression of tissue-specific genes in tumors gives a sensitive and robust diagnostic tool (10). In addition, the mis-expressed genes may produce immunogenic proteins, and render the tumor cells amenable to immunotherapy (11,12). Many of the tissue-restricted genes that are illegitimately

*To whom correspondence should be addressed. Tel: +33 1 57 27 89 16; Fax: +33 1 57 27 89 11; Email: pierre-antoine.defossez@univ-paris-diderot.fr
Correspondence may also be addressed to Olivier Kirsh. Email: Olivier.Kirsh@univ-paris-diderot.fr

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

re-expressed in tumor cells are normally only expressed in the testis; these genes are called Cancer/Testis (C/T) genes (13). However, other tissue-restricted genes, and in particular placental genes, may also be reactivated in tumors (10).

The goal of the present work was to identify chromatin regulators and signaling kinases which could be involved in illegitimate gene expression, to determine the interconnection between these molecular actors, and to test the physiological relevance of these findings.

Using high-throughput unbiased approaches, we report that most tissue-restricted genes examined are remarkably resistant to reactivation by a single hit in signaling pathways or chromatin regulators, suggesting that their reactivation in cancer results from a combination of events occurring during transformation.

An exception to this rule is the developmental gene ADAM12, highly expressed in the placenta, which encodes a metalloprotease re-expressed in cancers of diverse origins, such as breast, lung, liver, and colon malignancies (14–18). The oncogenic role of ADAM12 is especially clear in the case of Triple-Negative Breast Cancer (19).

We find that ADAM12 can be robustly induced in normal lung cells by stimulating MAP3K7/TAK, a kinase in the non-canonical TGF- β signaling pathway (20). This provides a mechanism for the known responsiveness of ADAM12 to TGF- β in cancer cells (21–25). ADAM12 can also be induced by depleting the histone deacetylase SIRT6 or the histone acetyltransferase GCN5/KAT2A. This repressive role of KAT2A is unusual, and we explain it by showing that KAT2A acts upstream of TAK1 and interacts with TAK1. Finally, our bioinformatic analyses argue that these mechanisms are physiologically relevant in the context of human cancer.

These data show that TAK1 inhibition by existing, well-tolerated drugs, could be an avenue to prevent illegitimate ADAM12 induction and decrease transformed phenotypes in several cancer types. More broadly, they describe unexpected connections between signaling pathways and chromatin regulators, and they reveal rules underpinning tissue-specific gene regulation in normal cells and tumors.

MATERIALS AND METHODS

Reagents and antibodies

The following antibodies were employed in this study: mouse ADAM12 (Proteintech 14139-1-AP); human ADAM12 (Sigma HPA030867); human TAK1 (SCBT sc-1839); human KAT2A (SCBT sc-20698); human SIRT6 (Abcam ab62739); human SMAD3 (ab28379), human phospho-SMAD3 (Abcam ab52903), human tubulin (Abcam ab7291), human TAB1 (CST 3226); human Histone H3 (CST 2650). TGF- β was from Proteintech and the TAK1 inhibitor (5Z)-7-oxozeaenol from Sigma.

Cell culture

MRC5, IMR90, SW39, SUM159PT, MDA-MB-231 and HEK293T were cultured in DMEM medium supplemented with 10% FBS and 1% penicillin/Streptomycin. BT549 cells were cultured in RPMI 1640 medium supplemented with 10% FBS and 1% penicillin/streptomycin. All the cell lines

were cultured in a humidified atmosphere at 37°C under 5% CO₂. The identity of all the cell lines was verified using the Eurofins cell line authentication service. We are grateful to Annabelle Decottignies and Woodring Wright for the gift of SW39 cells.

Virus production

The production of retroviruses and infection by the library was done as described in our previous publication (26). The plasmids expressing myristoylated kinases were a gift from William Hahn and Jean Zhao (Addgene kit # 1000000012; list of the kinases in Supplementary Table S1). All the plasmids were grown and prepared individually; twenty randomly selected vectors were sequenced and all contained the expected insert. IMR90 were seeded into six-well plates, infected, and total RNA was recovered 4 days after infection.

siRNA screening

For the siRNA screen seventy thousand IMR90 cells were seeded per well in six-well plates, reverse transfected with siRNAs using Dharmafect 1 (Dharmacon, Horizon Discovery), and total RNA was extracted 5 days after transfection. The protocol is described in detail elsewhere (27). For the screen we used ON-TARGETplus siRNA SMART-pools (Dharmacon, GE Healthcare), as well as control non-targeting pools (list of genes targeted in Supplementary Table S4). Additional siRNAs against TAK1, SIRT6 and KAT2A were purchased from Dharmacon and Sigma (sequences in Supplementary Table S6).

Selection of the genes of interest

We included 42 genes in the expression analysis (in addition to normalizers and other controls). These 42 genes were chosen according to several criteria. We first selected 20 tissue-specific genes inappropriately re-expressed in cancer from the data of Rousseaux and colleagues (10). We chose ten genes for which re-expression correlates with loss of DNA methylation as judged by 450K array data (MAGEB6, BRDT, DPEP3, RNF17, DDX4, SPATA22, TPTE, TUBA3C, DAZL, C10orf82), and 10 genes for which there is no such correlation (ADAM2, ADAM12, ADCY10, ASZ1, C9orf11, ALAS1, DDX53, ATAD2, RFX4, HORMAD1). We selected five 5-aza-cytidine inducible cancer/testis genes (MAGEA3, MAGEA4, MAGEA12, NY-ESO-1/CTAG1B and TKTL1) from a different publication (28). The rest of the genes were chosen from literature searches.

Nanostring analysis

The analysis was done by the Genomics Platform of Institut Curie (Paris, France). RNAs were analyzed with the BioAnalyzer using a Nano LabChip to assess their integrity (Bioanalyzer 2100 RNA 6000 Nano Kit From Agilent Technologies) and with a Nanodrop (Thermo) to assess their purity and concentration. RNA abundance was then measured with Nanostring technology (Nanostring Flex nCounter analysis system). All RNA processed to analysis

displayed a RIN >7.6 and a 260 nm/280 nm Ratio >1.8. Raw counts were first normalized with internal controls (Nanostring POS controls) and with the expression of three housekeeping genes (PGK1, TBP and TUBB2A). Nanostring probes are listed in Supplementary Table S2. The probes were tested against 'Human Reference Total RNA' (Agilent #750500), a mixture of mRNA from 10 human cancer cell lines (breast adenocarcinoma, cervix adenocarcinoma, hepatoblastoma, glioblastoma, melanoma, liposarcoma, histiocytic lymphoma, T lymphoblastic leukemia, plasmacytoma, and testicular embryonal carcinoma).

Statistical analysis of Nanostring data

For each probe, we plotted the distribution of log₂ normalized counts for all samples (controls and experiments). In the figures, we indicate two thresholds: mean + 2.5 standard-deviations, and mean + 3.5 standard-deviation. For normal distributions, these values correspond to a *P*-value of 1% and 0.05% after a two-tailed *t*-test, respectively. A Shapiro test showed that not all distributions were normal, so we did not indicate *P*-values in the figures.

Quantitative real-time PCR

RNA extraction was done using Tri reagent according to the manufacturer's recommendations. RNA was DNase treated, reverse transcribed using Superscript III (Invitrogen) and Oligo dT primers. qPCR was performed using Power SYBR Green (Applied Biosystems) on a ViiA 7 Real-Time PCR System (LifeTech), as described previously (29). TBP and PGK1 genes were used for normalization of expression values. The sequence of qRT-PCR is given in Supplementary Table S5.

Immunoblotting

Cells were harvested and lysed in RIPA buffer with protease and phosphatase inhibitors, sonicated (series of 30 s ON, 30 s OFF during for 5–10 min; Bioruptor, Diagenode). Protein extract was reduced with NuPage sample reducing agent and LDS sample buffer (LifeTech) as previously described (30). Fifty micrograms of protein was loaded for each sample.

Concanavalin enrichment (ADAM12 western blots)

Cells were harvested and lysed in RIPA buffer with protease and phosphatase inhibitors, sonicated (series of 30 s ON, 30 s OFF during for 5–10 min; Bioruptor, Diagenode). Protein extract from 200 and 400 µg of protein was incubated overnight with concanavalin A beads. The beads were washed five times with RIPA buffer and LDS sample buffer and sample reducing agent was added to the beads. The beads were boiled during for 5 min at 95°C and resolved by SDS–polyacrylamide gel electrophoresis. Proteins separated by electrophoresis were and electroblotted onto a nitrocellulose membrane using standard protocols (31).

Treatment of mice with (5Z)-7-oxozeaenol

M12CIG mice (32) were injected intraperitoneally with TAK1 inhibitor (5 mg/kg) one day prior to cardiotoxin injury and every day subsequent to injury. Cardiotoxin injury was performed as follows: anesthetized mice were injected with 50 µl of 10 µM cardiotoxin (Latoxan) in tibialis anterior muscles. Three days after injury the mice were sacrificed and the muscles were dissected, with half of each muscle used for protein extraction and the other half fixed with paraformaldehyde for immunofluorescence. The sections were stained with anti-GFP antibodies (Thermo Fisher #A11122, 1:1000) and counterstained with Alexa Fluor 488 anti-rabbit antibody (Invitrogen).

Co-Immunoprecipitation

Total protein extract was prepared by mixing cells with lysis buffer as previously described (33,34). The extract was incubated overnight in a cold room with agitation in the presence of 2 µg TAK1 antibody pre-incubated with magnetic beads coupled to protein G (Invitrogen) for 2 h. The beads were then washed seven times with 1 ml of wash buffer. The adsorbed proteins were dissociated by boiling beads for 10 min in 24 µl of Laemmli buffer and resolved by SDS-polyacrylamide gel electrophoresis. Proteins were separated by electrophoresis and electroblotted onto a nitrocellulose filter as previously described (35).

Wound healing assay

SUM159PT cells were plated in 12-well dishes and grown to confluence. During this time, cells were treated with TAK1 inhibitor or vehicle (DMSO). Then a scratch was performed in the middle of the well using a P10 plastic tip and image acquisition was performed every two hours using the Incucyte system.

TCGA analysis

TCGA gene count datasets for lung, colon and breast normal and cancer samples were downloaded from *Recount2* (36). ADAM12 expression normalized with *DESeq2* (37) was used to stratify tumors. We selected subgroups of tumors belonging to the first decile and the last decile of ADAM12 expression, and we compared the expression of TAK1, KAT2A and SIRT6 in these groups relative to the respective healthy tissue.

TAK1 signature characterization, TAK1 activation score, and correlation analysis

The TAK1 signature gene set was defined using public microarray data (GSE65069), re-analyzed with the *LIMMA* package (38). Using this data set, we first identified 516 genes that respond to TGF-β with a Fold-Change >1.5 and a *P*-value <0.05. Then we identified the 190 genes that lost induction by TGF-β in the presence of the TAK1 inhibitor (5Z)-7-oxozeaenol. This list is called the 'TAK1 signature' and is presented in Supplementary Table S7.

For correlations, we compute a 'TAK1-activation score', as the sum of the fold changes of the 190 genes in the TAK1 signature, in a particular tumor relative to normal tissue.

Lastly, we computed the Pearson correlation between this TAK-activation score and *ADAM12* or *KAT2A* expression. To test the significance of this correlation, we performed 1000 drawings of 190 random genes, and calculated the Pearson correlation between each random signature and the expression of *ADAM12* or *KAT2A*. The distribution of correlation values was plotted, and the *P*-value of the correlation for the TAK1 signature was estimated using the Gumbel approximation.

RESULTS

Most tissue-restricted genes are not illegitimately induced by a single hit in signaling pathways

We set out to determine whether alteration in signaling pathways, or alteration of chromatin factors, could reverse the silencing of tissue-restricted genes in otherwise normal cells. For this we developed two genetic screens, using as a model the non-transformed human lung cells IMR90 (Figure 1A).

We first sought to identify signaling pathways that, when activated, could turn on tissue-restricted gene expression (Figure 1B). For this, we used a collection of 192 kinases (list in Supplementary Table S1), genetically activated by the addition of a myristoylation signal, and encoded by retroviral vectors (39). IMR90 cells were grown in wells, and each well was infected with an individual kinase; in addition some wells were infected with the matching empty retroviral vector (control). After infection we extracted total RNA and used Nanostring probes to measure: the expression of the activated kinases (using the common sequence encoding the myristoylation-Flag tag); the abundance of spike-in positive and negative controls; the expression of four housekeeping genes for normalization; and the expression of a diverse set of 42 tissue-restricted genes (list of probes in Supplementary Table S2). These genes are expressed weakly or not at all in IMR90, and were chosen to reflect the distribution of illegitimately expressed genes in cancer, with a majority of testis-specific genes, some placenta-specific genes, and a few other cases (list and characteristic of genes in Supplementary Table S3, criteria for selection of the genes explained in the material and methods sections). The 42 genes under study were also chosen to represent potentially different mechanisms of epigenetic silencing. For instance, 23 of the genes contain a CpG island, and of those 18 are partially or fully methylated in IMR90 (Supplementary Table S3). The 19 genes that do not contain a CpG island are likely repressed by a DNA methylation-independent mechanism.

We included in the sample series, as positive controls, RNA from transformed IMR90 derivatives (SW39 cells), and a mixed RNA sample prepared from 10 different human cancer cell lines (Breast adenocarcinoma; cervix adenocarcinoma; hepatoblastoma; glioblastoma; melanoma; liposarcoma; histiocytic lymphoma; T lymphoblastic leukemia; plasmacytoma; and testicular embryonal carcinoma). We observed that 38 of the 42 tissue-restricted gene probes showed 2-fold or more increased signal relative to control in SW39 or in mixed human cancer mRNA, showing their potential to detect an increased signal (Supplementary Figure S1A).

We also measured the expression of the myristoylated kinases (using the common sequence encoding the myristoylation-Flag tag); all were detected, whereas non-infected cells IMR90 or SW39) showed no expression of the tag (Supplementary Figure S1B).

Having thus validated the expression of the kinases, and the ability of the Nanostring probes to detect an induction in gene expression, we next analyzed the expression of the 42 tissue-restricted genes in the ~200 samples (control and infected cells). The normalized data are presented in Figure 1C. They show, strikingly, that very few of the ~8400 probe/kinase pairs tested show a *Z*-score >2.5 (this corresponds to *P* = 1% assuming the data distribution is normal). One notable outlier is the gene *ADAM12*, which was strongly induced by the kinase MAP3K7/TAK1 (*Z*-score = 8.47, *P*-value < 10⁻⁵).

Most tissue-restricted genes are not illegitimately induced after depletion of a single chromatin factor

The second genetic screen we carried out aimed at determining whether the depletion of chromatin regulators (by RNAi) could be sufficient to reactivate tissue-restricted genes in otherwise normal cells. A further goal of this screen was to determine whether the genes reactivated by signaling pathway alterations were also sensitive to the depletion of chromatin factors, which would open the possibility of mechanistic studies to link signaling and chromatin.

We used a custom siRNA library targeting 160 chromatin factors (such as methyl-CpG-binding proteins, DNA modifying enzymes, lysine acetyltransferases and deacetylases, methyltransferases and demethylases; full list in Supplementary Table S4). As above, IMR90 cells were grown in wells, each well transfected with an siRNA pool targeting one specific factor, total RNA was extracted and tested by Nanostring analysis (Figure 1D). To evaluate the efficiency of the RNAi approach, we included in our Nanostring measurements 8 genes that were targeted by one of the siRNA pools (DNMT1, PPM1D, TET2, TET3, UHRF1, ZBTB4, ZBTB33, ZBTB38, Supplementary Figure S1C). Upon transfection of the specific siRNA, we observed a moderate depletion (~2-fold) for two of the controls (TET2 and TET3), and a greater than 5-fold mRNA depletion for the remaining six controls (DNMT1, PPM1D, UHRF1, ZBTB4, ZBTB33, ZBTB38).

Having validated the approach, we next analyzed the expression of the 42 tissue-restricted genes in the ~160 samples (control and transfected cells). The normalized data, presented in Figure 1E, show that only 21 of the ~6700 probe/siRNA pairs tested show a *Z*-score >2.5.

For 25 genes out of 42, no RNAi gave an induction of *Z* >2.5. Thirteen genes out of 42 had a single hit with *Z* >2.5. Finally, just four genes out of 42 had two hits with *Z* >2.5: *ASZ1*, *DDX53*, *FMR1NB* and *ADAM12*. *ADAM12* was induced by depletion of the deacetylase *SIRT6* (*Z*-score = 3.12, *P*-value = 9 × 10⁻⁴), and also by depletion of the acetyltransferase *KAT2A* (*Z*-score = 2.53, *P*-value = 5.7 × 10⁻³). Neither *SIRT6* nor *KAT2A* were able to activate any of the other tissue-restricted genes.

The results of this screen therefore show that most tissue-restricted genes are refractory to induction by depletion of a

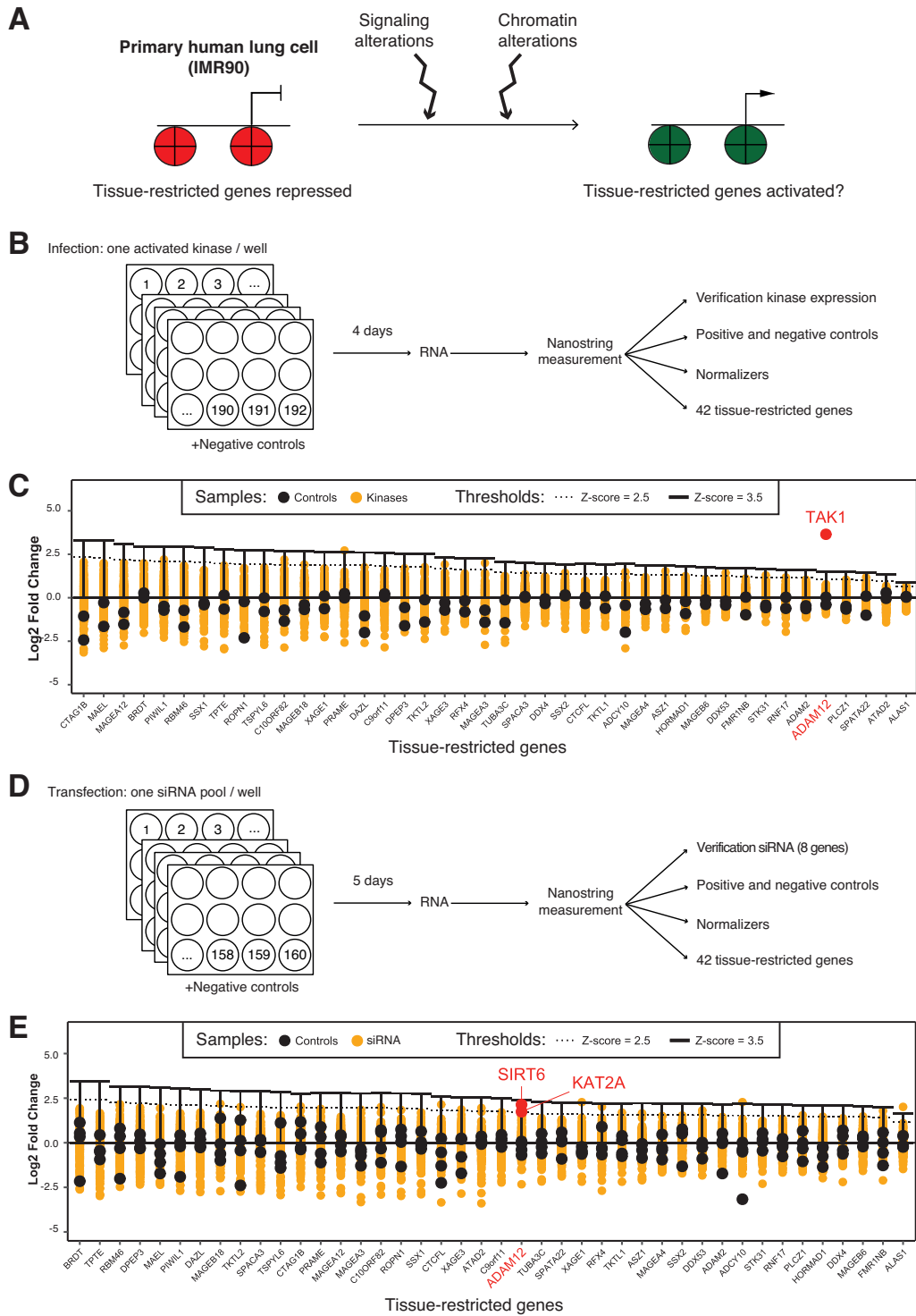


Figure 1. Two screens to investigate the mechanisms controlling tissue-restricted gene reactivation converge on ADAM12. (A) Principle of the approach. Primary human cells were challenged by alterations in signaling pathways or chromatin regulators, and we determined whether these changes were sufficient to cause inappropriate expression of tissue-restricted genes. Red is used to represent repressive chromatin, and green permissive chromatin. (B) Design of the signaling screen: the primary cells were infected with 192 different genetically activated kinases, and gene expression assayed by Nanostring. (C) Results of the signaling screen: \log_2 fold change of expression for each of the 42 tissue-restricted genes following infection by the 192 activated kinases. Each infection is represented by a yellow dot. The negative controls are indicated as black dots. The only kinase/gene pair showing a Z-score >3.5 is the gene ADAM12 being highly activated by the kinase MAP3K7/TAK1 (red dot). (D) Design of the chromatin screen: the primary cells were transfected with 160 different siRNA pools, each targeting a specific chromatin regulator, then gene expression was assayed by Nanostring. (E) Results of the chromatin screen: \log_2 fold change of expression for each of the 42 tissue-restricted genes following transfection of the 160 siRNA pools. Each transfected sample is shown as a yellow dot, controls are shown as black dots. Only a few siRNA/gene pairs have a Z-score >2.5 . These include the induction of ADAM12 by depletion of SIRT6 and KAT2A.

single chromatin factor. Again, one exception to this rule is the gene ADAM12, for which expression is induced in both screens. We focused the rest of the work on this gene, in order to validate our findings, study their physiological relevance in the context of cancer, and clarify the interplay of signaling factors and chromatin regulators in its transcriptional regulation.

The kinase activity of TAK1 is necessary for the induction of ADAM12

As shown in Figure 2A, expression of ADAM12 was induced only by one kinase: TAK1. The ADAM12 mRNA induction was verified by RT-qPCR after an independent infection with the TAK1 expressing vector (Figure 2B); in contrast a point mutant of TAK1 devoid of kinase activity (TAK1 catalytic dead, or TAK1 CD), failed to induce ADAM12 expression. Wild-type TAK1, but not the CD mutant, also induced the ADAM12 protein, as shown by western blotting (Figure 2C). Similar results were also obtained in an other non-transformed human lung cell line, MRC5 (Supplementary Figure S2A).

TAK1 links TGF- β to ADAM12 induction

TAK1 is activated by TGF- β (20), and more specifically it acts in the non-canonical branch of the TGF- β pathway, stimulating the transcription of genes such as IL-6 (Figure 2D). This non-canonical branch is distinct from the canonical branch, which involves SMADs, and leads to activation of genes such as CDKN2B (Figure 2D). It has also been shown that ADAM12 can be induced by TGF- β (21–25). However, to the best of our knowledge, the kinase mediating this activation has not been reported. Our previous results suggested that TAK1 could be a candidate.

To test this hypothesis, we first used a well-characterized chemical inhibitor of TAK1, (5Z)-7-oxozeaenol (40), referred to as '5Z' hereafter. We treated IMR90 cells with TGF- β , in combination with 5Z or just its solvent, DMSO. As expected, TGF- β treatment (5 ng/ml, 6 h), induced all the target genes we examined (ADAM12, IL-6, SMAD7, CDKN2B; Figure 2E). Combining 5Z with TGF- β blocked the induction of ADAM12 and IL-6, whereas it did not affect CDKN2B induction, which was in fact potentiated. We also verified that 5Z treatment did not interfere with SMAD3 phosphorylation, a readout of the canonical pathway activation (Figure 2F); furthermore similar results were obtained in MRC5 cells (Supplementary Figures S2B and C).

The data obtained in cell culture strongly suggested that TAK1 mediated the induction of ADAM12 by TGF- β . To assess whether this was also the case in an *in vivo* setting, we used transgenic reporter mice in which GFP is driven by the ADAM12 promoter (Figure 2G). In these animals, wounding the muscle with cardiotoxin leads to the formation of ADAM12-positive myofibroblasts, which are marked by GFP, and this process depends on TGF- β (32). Treating mice with 5Z before and after injury led to a significant decrease of GFP+ cells in the regenerating muscle (Figures 2H and I). Consistent with a decrease in the number of GFP+ cells, the ADAM12 protein was also decreased

in muscle lysates, as measured by western blot, when TAK1 was inhibited (Supplementary Figures S2D and E).

Altogether these data show that TAK1 activation can induce ADAM12 expression in normal cells. In addition, TAK1 activity is necessary to induce ADAM12 expression *in vitro* and *in vivo*, in a mouse injury model.

An RNAi screen identifies chromatin modifiers that regulate ADAM12 expression

The top two siRNA targets causing reactivation of ADAM12 were SIRT6 and KAT2A (formerly known as GCN5, Figure 3A). The results were validated with an additional siRNA not present in the initial pool (Figures 3B and C), and similar observations were made in MRC5 cells (Supplementary Figures S3A–C). Notably, combined knockdown of the two genes led to higher ADAM12 induction than either individual knockdown, suggesting they might act in different pathways (Figure 3B).

SIRT6 is a well-known transcriptional repressor that acts in part by deacetylating histones at H3K9 (41). In contrast, KAT2A generally mediates transcriptional activation, by acetylating lysines, including H3K9, as part of the SAGA complex (42). Therefore, it was unexpected that knocking down KAT2A, an activator, led to ADAM12 induction, and thus we chose to investigate further this surprising finding.

KAT2A acts upstream of TAK1 and interacts with TAK1

KAT2A is known to act on certain non-histone proteins, such as the kinase PLK4 (43). Therefore we considered the hypothesis that KAT2A might act upstream of TAK1, and negatively control its activity, which could account for the induction of ADAM12 upon KAT2A knockdown.

To test this idea, we first performed TAK1 knockdown simultaneously with KAT2A knockdown. In that situation, depletion of KAT2A failed to induce ADAM12 (Figures 4A and B). In an independent approach, we also pre-treated cells with 5Z and then performed KAT2A knockdown; this also dampened ADAM12 induction (Supplementary Figures S4A and B).

The result of these two experiments shows that removal of KAT2A does not induce ADAM12 when TAK1 is absent or inactive. This result is compatible with the hypothesis that KAT2A functions upstream of TAK1 and negatively regulates its activity. As KAT2A can physically interact with certain non-histone proteins, we asked whether it might interact with TAK1. Immunoprecipitation of endogenous TAK1 was performed on MRC5 cells using the known co-factor TAB1 as a positive control (44). We also detected KAT2A in TAK1 immuno-precipitates (Figure 4C). The interaction pre-existed before TGF- β addition, and persisted after TGF- β addition (Figure 4C).

TAK1 drives ADAM12 expression in a triple-negative breast cancer cell line

After establishing that TAK1 mediates ADAM12 induction by TGF- β in normal human cells, we asked whether it was also involved in the sustained expression of ADAM12 seen in tumor cells, and particularly in Triple-Negative Breast Cancer (TNBC) cells.

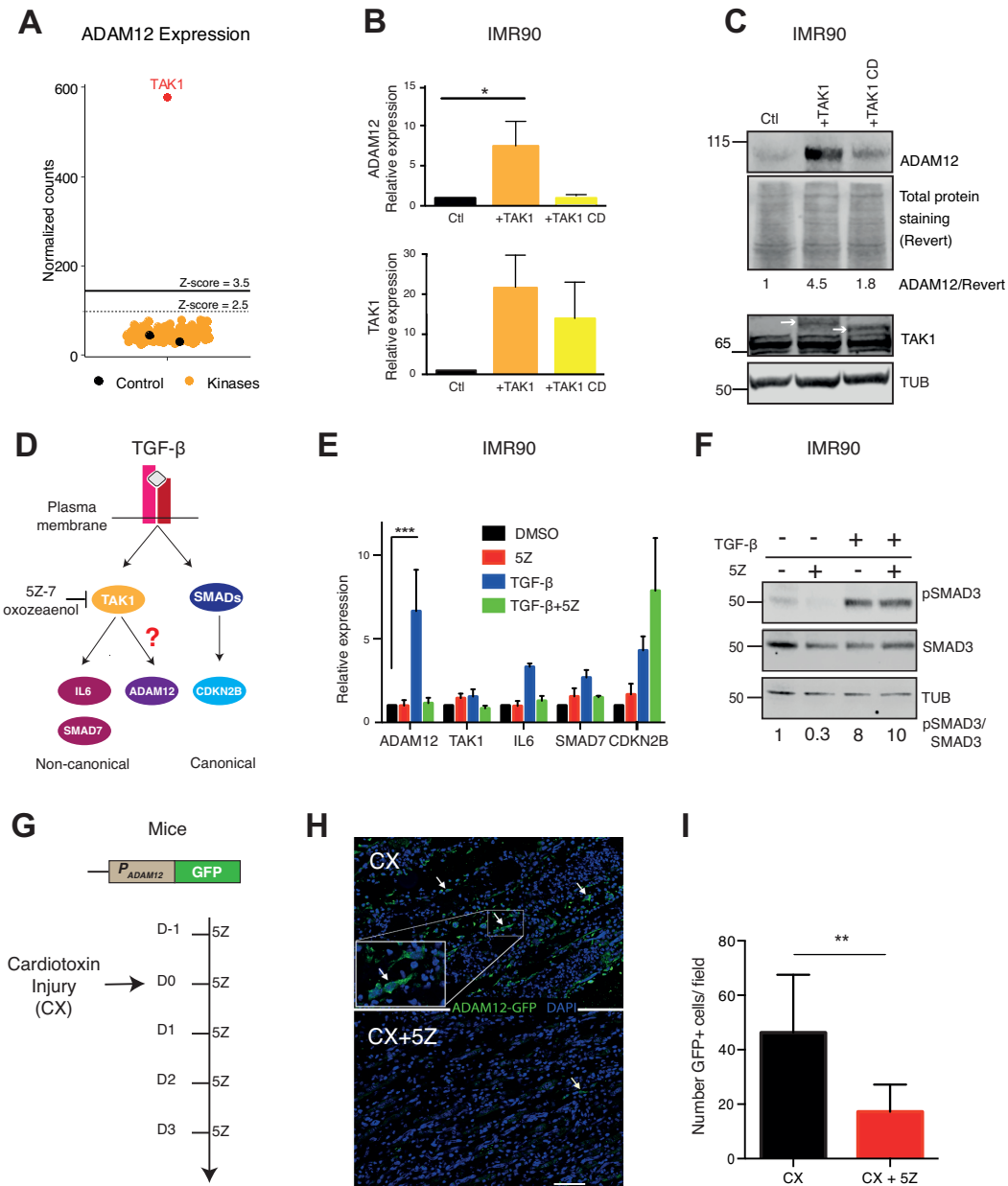


Figure 2. TAK1 mediates ADAM12 induction in vitro and *in vivo*. (A) Results of the signaling screen: only TAK1 expression resulted in the activation of ADAM12. Yellow and red dots represent ADAM12 normalized Nanostring counts in IMR90 infected with the kinase, each dot representing one kinase. Black dots represent ADAM12 expression level in control cells (infected with empty vector). (B) RT-qPCR validation of ADAM12 reactivation by TAK1 in IMR90 cells. ADAM12 was activated by the wild-type form of TAK1 but not by the catalytic dead form of TAK1 (TAK1 CD). (C) Western blot analysis of ADAM12 protein levels following the overexpression of wild-type and catalytic dead TAK1 (TAK1CD). Ctl: cells infected with the empty vector. The white arrows represent the exogenous TAK1; TAK1CD is lower because it does not self-phosphorylate. Tubulin (TUB) is the loading control. For ADAM12, concanavalin A enrichment was performed. (D) Schematic representation of canonical and non-canonical TGF- β pathways. TAK1 is a component of the non-canonical TGF- β pathway. (E) RT-qPCR on the indicated genes in the presence or absence of the TAK1 kinase inhibitor (5Z)-7-Oxozeaenol (5Z). IMR90 were pre-treated with 0.3 μ M 5Z or DMSO for two hours, followed by stimulation with 5 ng/ml of TGF- β for 6 h. ADAM12 induction by TGF- β is abolished by 5Z. (F) Control western blot showing the phosphorylation of SMAD3, indicating the activation of canonical TGF- β pathway even though TAK1 was inhibited by 5Z. (G) Testing the dependence of ADAM12 on TAK1 *in vivo*: reporter mice with the ADAM12 promoter (P_{ADAM12}) driving GFP expression were subjected to an injury (cardiotoxin injection into the tibialis muscle). Some of the mice were treated by intraperitoneal injection of TAK1 inhibitor (5Z) at 5 mg/kg before and after injury, while the controls were injected only with solvent (DMSO). Three days after the injury, mice were sacrificed and muscle tissue was dissected. (H) Immunofluorescence on injured muscle shows that 5Z treatment reduces GFP induction. The arrows indicate GFP-positive cells. Scale bar: 150 μ m. (I) Quantification of panel H (10 fields counted in each condition). All the experiments were performed at least three time except for ADAM12 western blot and mice experiments. The statistical analysis was performed with one way ANOVA followed by Dunnett's test, except for panel I where a Mann-Whitney test was performed. In all figures, we used the following conventions: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

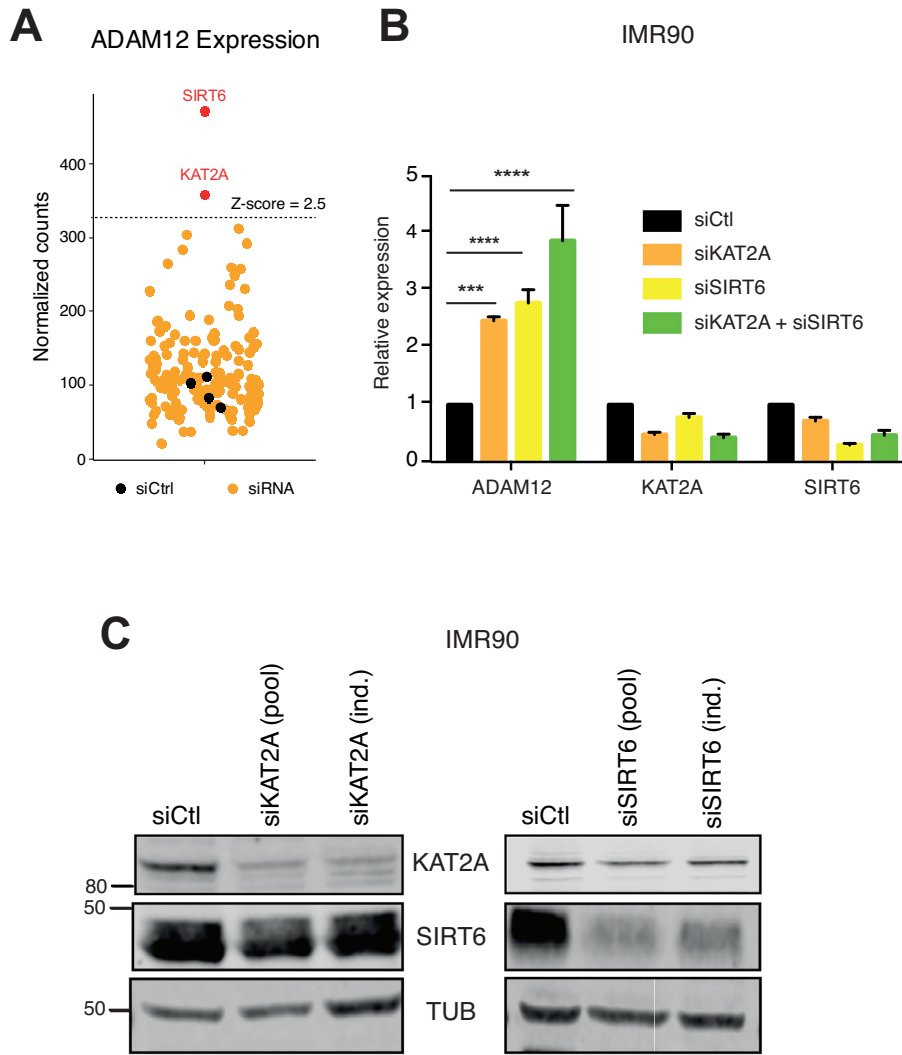


Figure 3. KAT2A and SIRT6 repress ADAM12 expression in normal cells. (A) Results of the chromatin screen: knockdown of KAT2A or SIRT6 causes the reactivation of ADAM12. Yellow dots represent ADAM12 normalized Nanostring counts in IMR90 transfected with targeting siRNAs; black dots for cells transfected with a non-targeting siRNA (siCtrl). (B) RT-qPCR representing relative expression of ADAM12, KAT2A and SIRT6 in IMR90 cells infected with siCtrl (non-targeting siRNA), siKAT2A and/or siSIRT6. These siRNAs are independent from those used in the screen. Statistical analysis was performed with a one-way ANOVA followed by a Dunnett's test. (C) Western blot analysis showing efficient down regulation of KAT2A and SIRT6 after transfection of either the original pool of siRNA used in the screen (pool), or of individual siRNAs (ind.).

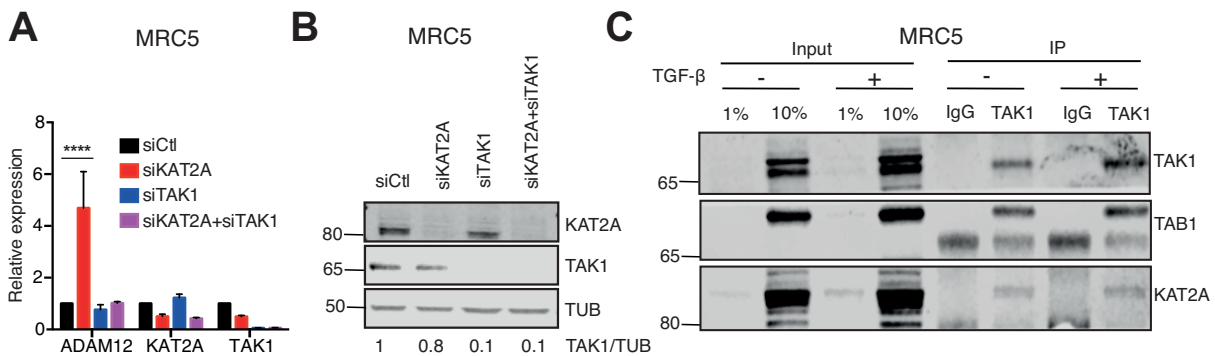


Figure 4. Epistasis and interaction between KAT2A and TAK1. (A) RT-qPCR on the indicated genes in MRC5 cells transfected with siCtrl (non-targeting siRNA), siKAT2A and/or siTAK1. ADAM12 is induced by the knockdown of KAT2A but this is abolished by the simultaneous knockdown of TAK1. Statistical analysis was performed with a two-way ANOVA followed by a Dunnett's test. (B) Western blot showing expression level of TAK1 and KAT2A after siRNA transfection. (C) Co-immunoprecipitation of endogenous TAK1 from MRC5 cells with or without stimulation by 5 ng/ml TGF- β revealed interaction with KAT2A. TAB1, a known interactor of TAK1, served as a positive control.

For this we used SUM159PT cells, which express ADAM12 (19). We first inhibited TAK1 chemically, using 5Z; this led to a decrease of ADAM12 mRNA and protein expression (Figures 5A and B), an effect also seen for another TNBC line, BT549 (Supplementary Figures S5A and B). We also knocked down TAK1 using two independent siRNAs in SUM159PT; this resulted in decreased ADAM12 expression as well (Figure 5C and D). Finally we found that, as in non-transformed cells, TAK1 interacts with KAT2A in SUM159PT cells (Supplementary Figure S5C).

The next question we sought to address was whether this molecular pathway could be shown to have phenotypical consequences. One of the known effects of ADAM12 is to promote cellular migration (16–18), so we tested this phenotype. We measured cell migration quantitatively, and found that treating SUM159PT cells with 5Z decreased their migration in a ‘scratch assay’ (Figure 5E). In parallel experiments, we quantified the effect of 5Z on the growth rate of cells after seeding in a non-scratched plate (Figure 5E). The growth of solvent-treated and 5Z-treated cells was identical, ruling out decreased proliferation as the cause of decreased migration.

KAT2A represses ADAM12 expression in a triple-negative breast cancer cell line

Next, we asked if our observation that KAT2A restricts ADAM12 expression, made in a non-transformed cell context, was also relevant in cancer. For this we used a different TNBC line, MDA-MB-231, in which the expression of ADAM12 is modest compared to SUM159PT (45). In this context as well, knockdown of KAT2A led to an increased level of ADAM12 mRNA (Figure 6A) and protein (Figure 6B). Importantly, this effect was not seen when TAK1 was simultaneously knocked down (Figures 6A–C). This establishes that the KAT2A/TAK1 regulatory axis is also functional in TNBC lines. Finally, we verified that the TAK1/KAT2A interaction was also detectable in MDA-MB-231 cells, both in the cytoplasmic and nuclear compartments (Supplementary Figure S6).

The analysis of human cancer expression data supports the regulation of ADAM12 by TAK1 and KAT2A in various tumor types

Lastly, we sought to determine whether our findings on normal or transformed cells represent a general mechanism. For this, we performed a bioinformatic analysis of TCGA data, starting with breast cancer.

We first verified the previously described observation that ADAM12 is upregulated in breast tumors (Figure 7A). We then stratified the breast tumor samples based on ADAM12 mRNA levels, and compared the high-ADAM12 group (highest decile for ADAM12 expression) to the low-ADAM12 group (lowest decile) and to normal tissue (Figure 7B). We found that low ADAM12 expression was associated with high KAT2A expression (Figure 7B). This association is compatible with the model we have put forward based on our mechanistic work with cell lines. Conversely, TAK1 was more expressed in high-ADAM12 tumors than

in low-ADAM12 tumors (Figure 7B), again consistent with our *in vitro* work. When the breast tumors were stratified by subtype, the analysis revealed that not only triple-negative tumors contained high-ADAM12 and low-TAK1 samples, but the other groups did as well (Figure 7C).

The mRNA level of TAK1 could be an imperfect proxy of its activity, thus to strengthen these data, we performed a more detailed analysis, which started by identifying a ‘TAK1 signature’ (Figure 7D). For this we used transcriptome data obtained in primary human cells, treated with TGF- β , 5Z or TGF- β +5Z (42). We identified 516 TGF- β responsive genes, which we divided into three classes. The first class contains 171 genes that are induced by TGF- β equally well in the presence or absence of 5Z and are therefore TAK1-independent. The second class contains 155 genes, including SMAD3, that are induced by TGF- β more strongly when 5Z is present; those may be repressed directly or indirectly by TAK1. The last class of genes are induced by TGF- β less strongly when 5Z is present, and are therefore potentially dependent on TAK1. For further analysis, we focused on this group of 190 genes (including ADAM12, SMAD7, and IL-6), referred to as a ‘TAK1 signature’ (Figure 7D and Supplementary Table S7).

To test its discriminative power relative to ADAM12 expression, we assembled a set of 338 TCGA breast samples (112 normal, 113 high-ADAM12, and 113 low-ADAM12), and performed unsupervised clustering based on the TAK1 signature. This resulted in near-perfect segregation of normal, ADAM12-low, and ADAM12-high breast tumors (Figure 7E). Therefore TAK1 activity is indeed strongly associated with ADAM12 expression in breast tumors.

This first analysis had been performed on a highly contrasted group of tumors (highest versus lowest ADAM12 expression), so we lastly asked whether it held true for the rest of the tumors as well. For this, we used as a metric the ‘TAK1 score’, which reflects the expression of the 190 genes in the TAK1 signature (see materials and methods).

Using all 1104 tumors in the dataset, we observed a clear correlation between the TAK1 activation score and the ADAM12 expression level (Pearson’s $r = 0.69$; Figure 7F). Conversely, the TAK1 activation score is significantly anti-correlated with the KAT2A expression level (Pearson’s $r = -0.45$; Figure 7G). A thousand random samplings failed to yield a set of 190 genes that displayed these correlations (Supplementary Figure S7A), supporting the validity of the TAK1 signature.

Finally, we sought to determine whether our findings may apply to cancer types other than breast malignancies. We repeated the same bioinformatic analyses on 1044 lung tumors (Supplementary Figures S7B–D), and on 505 colon adenocarcinomas (Supplementary Figures S7E–G). In both tumor types we found correlations between high ADAM12 expression and lower KAT2A levels (Supplementary Figures S7C and F). A high TAK1 signature very clearly separated high-ADAM12 from low-ADAM12 tumors both in lung (Supplementary Figure S7D) and colon (Supplementary Figure S7G), and the TAK1 activation score was positively correlated with ADAM12 expression, and negatively correlated with KAT2A expression, in both tumor types (Supplementary Figures S7H–I).

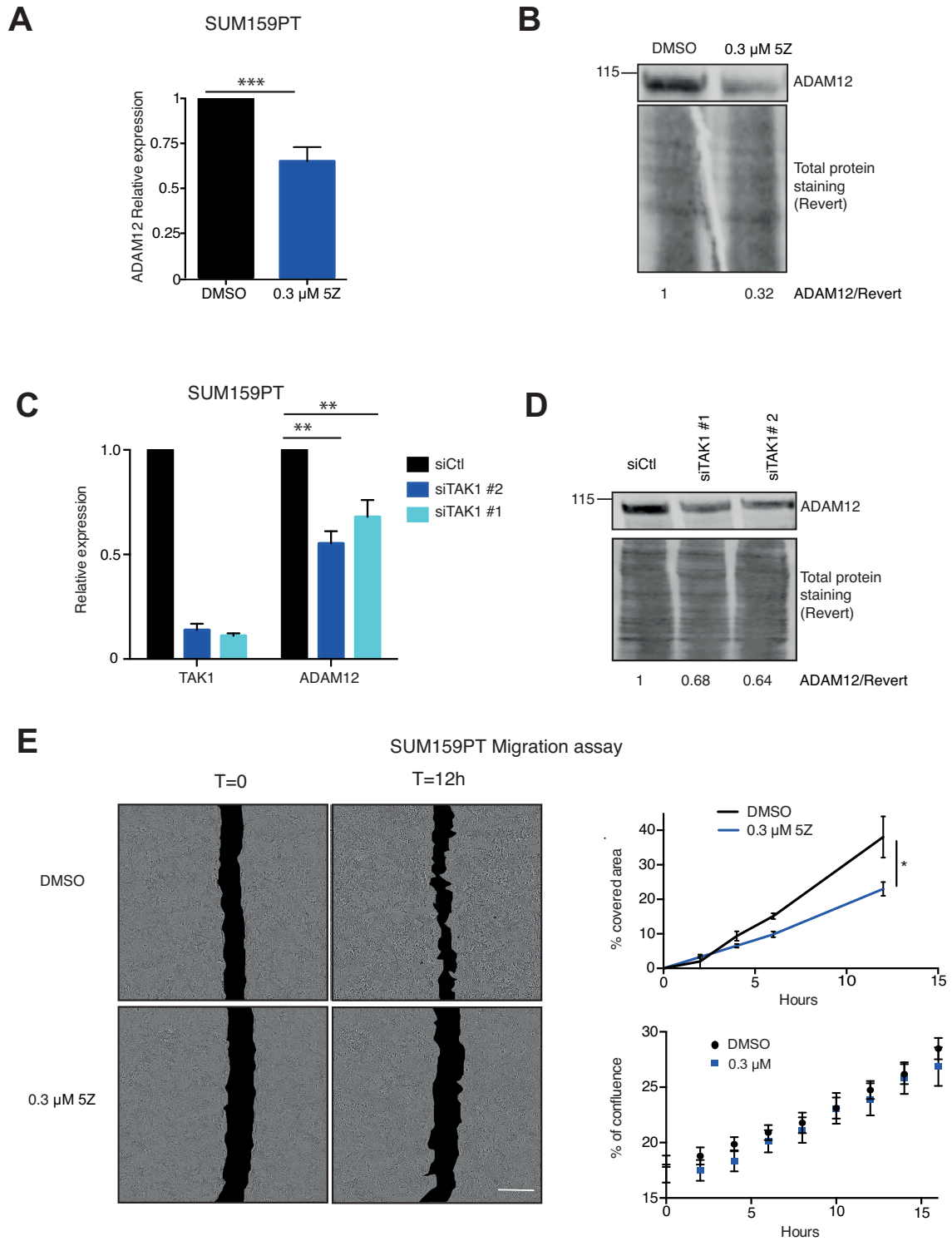


Figure 5. TAK1 is involved in ADAM12 expression in breast cancer cells. (A) SUM159PT breast cancer cells were treated with 0.3 μ M or 1 μ M 5Z for four days and the level of ADAM12 was assessed by RT-qPCR. Inhibition of TAK1 by 5Z reduced the level of ADAM12 transcripts. (B) Western blot analysis showing ADAM12 protein expression after 5Z treatment of SUM159PT cells. Concanavalin A enrichment was performed. (C) RT-qPCR showing the expression levels of TAK1 and ADAM12 in SUM159PT cells transfected with non-targeting control siRNA (siCtl) or siTAK1 for four days. Knockdown of TAK1 decreased the levels of ADAM12 transcripts. (D) Western blot showing ADAM12 expression after knockdown by siCtl or siTAK1 in SUM159PT cells. As in panel B. (E) Wound healing assay. SUM159PT cells were treated with 0.3 μ M and 1 μ M 5Z for 4 days, following treatment, a scratch was made in the dishes and it was imaged every 2 h using the Incucyte live cell system. The area of the wound at different time points was measured by ImageJ software, then a percentage of the area covered by time was determined and is plotted in the panel on the right. Inhibition by 5Z delays the wound healing process. The statistical analysis was performed employing a two way ANOVA test followed by a Dunnett's test. Bottom right panel: growth rate of the cells measured as their percentage of confluence after seeding in a non-scratched plate; 5Z does not affect the growth rate.

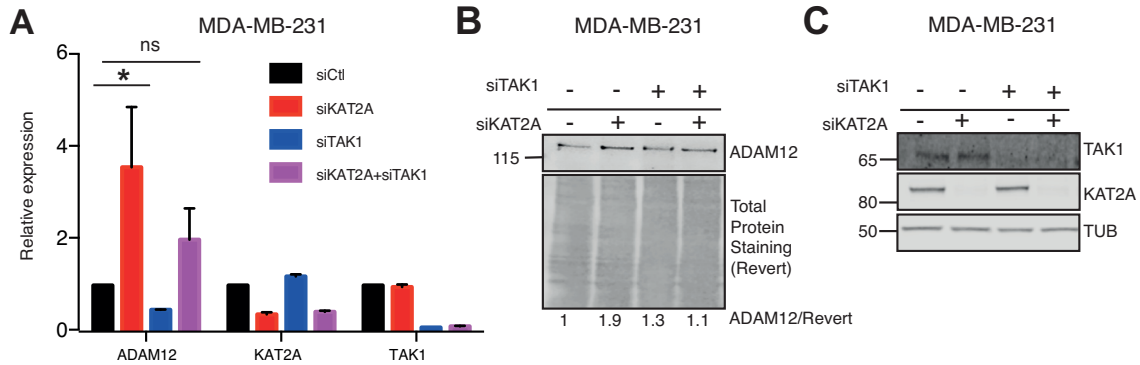


Figure 6. KAT2A negatively regulates ADAM12 expression in breast cancer cells. (A) RT-qPCR showing the relative expression level of ADAM12, TAK1 and KAT2A in MDA-MB-231 cells transfected with non-targeting control siRNA (siCtrl), siKAT2A, siTAK1 and/or siKAT2A for four days as depicted. The statistical analysis was performed with a two-way ANOVA test followed by a Dunnett's test. (B) Western blot showing ADAM12 expression after siRNA transfection in MDA-MB-231. For ADAM12, concanavalin A enrichment was performed. (C) Validation of siRNA efficiency by western blotting.

Altogether, these data agree with and extend our mechanistic data obtained *in vitro*, arguing that KAT2A negatively regulate ADAM12 expression in several prevalent human tumor types. We also conclude that TAK1 activity is a strong positive determinant of ADAM12 expression in lung, breast, and colon cancer.

DISCUSSION

Single hits affecting signaling pathways or chromatin regulation are generally not sufficient to activate tissue-restricted genes

We tested ~200 activated kinases for their capacity to induce tissue-restricted gene expression in non-transformed cells. While TAK1 could robustly activate ADAM12 in normal lung cells, we found no other instance of a tissue-restricted gene being strongly reactivated by a single activated kinase. The system we used has several possible limitations that need to be considered when interpreting this result. First, the kinases were genetically activated by the inclusion of a myristoylation signal, which causes recruitment of the kinases to the cell membrane. This could potentially lead to false negative results if a given kinase needs to be in a different cellular compartment, for instance the nucleus, to activate gene expression. Second, we have no formal proof that each single kinase was functional, but this collection has been validated in several previous publications (26,39,46–49), and we verified in our system that each kinase was expressed. We also verified the ability of the probes to detect an induction, even small.

Similarly, we tested 160 siRNA pools directed against chromatin modulators for their capacity to induce expression of the tissue-restricted genes. Again, most genes were refractory to this treatment, with one of the rare exceptions being ADAM12.

The tissue-restricted genes are embedded in repressive chromatin in non-expressing cells (10). For the illegitimate expression to take place, this chromatin has to become permissive, and it is possible that cell division helps this process, as chromatin is remodeled along with DNA replication (50). In the course of our genetic screens (4 or 5 days, see Figure 1), the IMR90 cells divide ~3 times, and we have

found that these conditions are appropriate to identify a negative regulator of a non-tissue restricted gene (27). However, we cannot rule out the possibility that longer incubations with the activated kinases or the siRNAs would lead to increased chromatin remodeling and additional tissue-restricted genes being reactivated.

We can conclude, however, that ADAM12 is more readily reactivatable than the other genes tested. This property does not seem to apply to all placenta-specific genes as XAGE3, which was also present in our gene set, was never induced in our screens. Therefore, ADAM12 likely undergoes a specific regulation.

ADAM12 is repressed by SIRT6: possible mechanisms

One of the negative regulators of ADAM12 expression identified in our siRNA screen is the histone deacetylase SIRT6. Interestingly, SIRT6 is often underexpressed in tumors, and its loss contributes to cellular transformation (51); a corresponding induction of ADAM12 could possibly contribute to this phenotype. Mechanistically, SIRT6, as several other sirtuins, has the capacity to repress gene promoters by causing histone deacetylation (41). Therefore a simple hypothesis is that SIRT6 acts directly on the ADAM12 promoter and keeps its expression low. ChIP-Seq has been performed on SIRT6 in mouse embryonic fibroblasts (52), and these data suggest that the ADAM12 promoter is directly bound by SIRT6, at least in this cell type. Other non-exclusive possibilities might also occur: SIRT6 could repress the expression of an ADAM12 activator. Finally, we note that SIRT6 has been shown to bind and activate KAT2A in certain contexts (53), therefore this might also contribute to ensuring the repression of ADAM12 by SIRT6.

TAK1 links TGF- β to ADAM12 induction: pathways and consequences

The kinase MAP3K7/TAK1 was a strong outlier in our signaling screen, as it potently activates the expression of ADAM12 even in lung cells. There have been several indications that TGF- β can induce ADAM12 expression (21–25). Now we bring evidence that the non-canonical branch

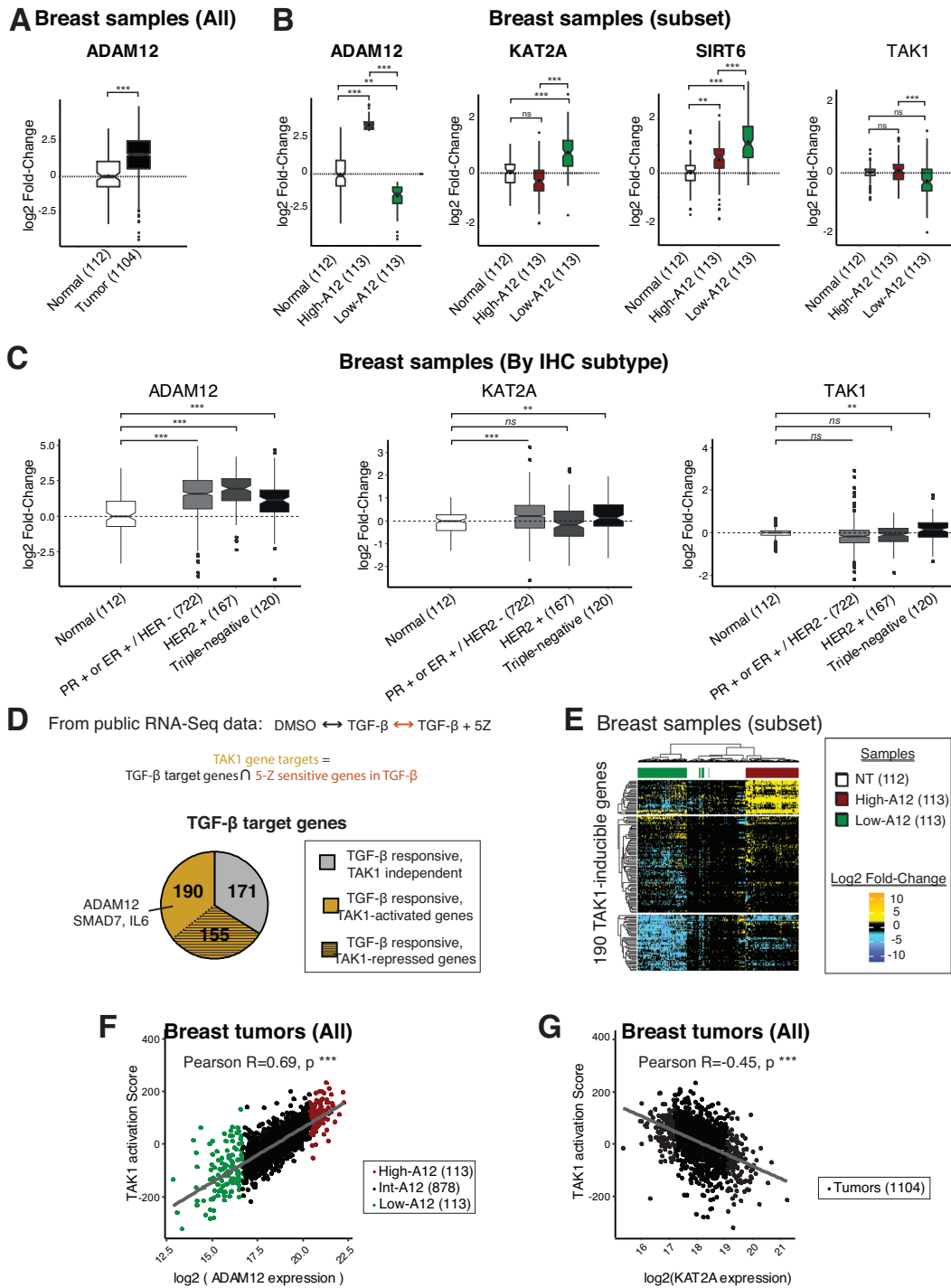


Figure 7. TCGA analysis revealed co-expression of ADAM12 with KAT2A, TAK1 and TAK1-inducible genes signature in breast tumors. (A) ADAM12 expression in breast tumors and normal breast tissue from the TCGA database. The boxplots contain 50% of the values, with a notch at the median value, and a diamond at the average value. The whiskers depict the first and last quartiles, and outliers appear as black dots. (B) ADAM12, KAT2A, and TAK1 expression in normal breast tissues and a subset of breast cancer samples from the TCGA database, selected based on their low (first decile, green) or high (ninth decile, red) ADAM12 expression. (C) ADAM12, KAT2A, and TAK1 expression in normal breast tissues and breast cancer samples from the TCGA database stratified by subtype. (D) Definition of the TAK1 signature, using transcriptomic data for primary cells treated with TGF- β with or without the TAK1 inhibitor 5Z. The TAK1 signature contains the genes that are upregulated by TGF- β addition, but not upregulated in the TGF- β +5Z condition. (E) Hierarchical clustering with euclidean distance metric and Ward's linkage method of highest and lowest ADAM12 expressing tumors samples, and normal breast tissue samples. The unsupervised clustering of breast tumors according to the TAK1 signature almost perfectly segregates tumors according to ADAM12 expression. (F) TAK1 activation score (calculated from the expression of genes in the TAK1 signature) correlates positively with ADAM12 expression in 1104 breast tumors. The high-ADAM12 samples are shown in red and low-ADAM12 in green. The statistical analysis was performed with a two-way ANOVA test, followed by a Tukey HSD test (***) denotes $P < 0.001$). (G) TAK1-activation score negatively correlates with KAT2A expression in 1104 breast tumors. Statistics as in panel E.

of TGF- β signaling is especially involved, via the involvement of TAK1. Several molecular pathways are activated downstream of TAK1 (54), one of which is NF- κ B (55–57). As NF- κ B has been shown to upregulate ADAM12 in response to TGF- β (24), it may constitute the next element in the signalling cascade linking TGF- β , TAK1 and ADAM12 induction.

ADAM12 has been shown to contribute positively to tumorigenesis (17,19,58), therefore its extinction is predicted to have beneficial effects. TAK1 inhibitors have been actively investigated in the context of immunity, fibrosis, and cancer (59–62). Our results suggest that these molecules could be particularly relevant in ADAM12-positive tumors. Based on our TCGA analysis, these represent a large number of potentially targetable cases.

A repressive role for KAT2A

We report that depletion of KAT2A increases TAK1 expression in multiple cellular contexts. This result is counterintuitive given that KAT2A, a core component of the SAGA complex, is a global activator of PolII transcription (63). Our experiments with chemical inhibitors and RNAi show that KAT2A depletion has no effect on ADAM12 expression when TAK1 is absent or inactive. The simplest interpretation of these findings is that KAT2A inhibits the activity of TAK1, and an indication that this might occur is our finding that KAT2A and TAK1 co-immunoprecipitate. Future work will determine whether KAT2A affects TAK1 protein abundance, localization, or activity. Such a mechanism has recently been demonstrated for the Polo-Like Kinase PLK4, the activity of which is inhibited by KAT2A-mediated lysine acetylation (43).

Illegitimate gene expression in cancer

Our work using normal cells show that most tissue-restricted genes are more resistant to activation signals than ADAM12. This may reflect the existence of several layers of epigenetic repression, for instance repressive histones coupled to DNA methylation, which would require more than one event to be relieved. Two predictions from this hypothesis are that ADAM12 should be reactivated in more tumors than the other tissue-restricted genes, and also that the expression of tissue-restricted genes may be a rather late event during oncogenesis, occurring only after several transforming events have accumulated. If this prediction is correct, then immunotherapy on illegitimately expressed genes would be of limited efficacy on tumors in their early stages.

DATA AVAILABILITY

The Nanostring data have been deposited in GEO under reference GSE124101.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thanks the Genomics Platform of Institut Curie for Nanostring analysis. The Nanostring platform of Institut Curie was initiated with the support of French grants (LabEx and EquipEx): ANR-10-IDEX-0001–02 PSL, ANR-11-LBX-0044 and « INCa-DGOS- 4654 » SIRIC11–002. We thank Valérie Doye, Stéphanie Bolhy, and Philippe Girard at Institut Jacques Monod for help with the Incucyte. Concerning the lentiviral vector production, we acknowledge Christophe Huret and the Paris Diderot Vectorology Platform of Paris Diderot University, Sorbonne Paris Cité, CNRS UMR 7216, Paris, France. We acknowledge the kind help of Elisabeth Brambilla with some experiments that do not appear in the submitted version of the manuscript.

For the gift of reagents we acknowledge the kind contribution of Katrin Chua, Philippe Chavrier, Annabelle Decotignies, Sébastien Jauliac and Michael Kracht.

Finally, we thank the following colleagues for useful discussions: Jean-Marc Vanacker, Robert Dante, Nathalie Théret, Saadi Khochbin, Céline Prunier, Céline Vallot, Claire Rougeulle, Isabelle Bernard-Pierrot, Slimane Ait-sali, Jonathan Weitzman, Valérie Lallemand-Mezger.

FUNDING

National Institutes of Health [EpiCALC 2013 to P.A.D. and D.B.]; Institut National du Cancer [PLBio15–013 to P.A.D. and D.B.]; Labex ‘Who am I?’ [ANR-11-LABX-0071 under ANR-11-IDEX-005–02, to P.A.D.]; European Research Council [ERC648428-PERIF to L.P.]; Fondation ARC [4th-year PhD scholarship to I.N., Programme Labellisé PGA1/RF20180206807 to P.A.D.]; Fondation pour la Recherche Médicale [postdoctoral fellowship to A.H.]. The open access publication charge for this paper has been waived by Oxford University Press - *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES

- Lee, T.I. and Young, R.A. (2013) Transcriptional regulation and its misregulation in disease. *Cell*, **152**, 1237–1251.
- Tee, W.-W. and Reinberg, D. (2014) Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development*, **141**, 2376–2390.
- Suganuma, T. and Workman, J.L. (2013) Chromatin and signaling. *Curr. Opin. Cell Biol.*, **25**, 322–326.
- Sen, P., Shah, P.P., Nativio, R. and Berger, S.L. (2016) Epigenetic Mechanisms of longevity and aging. *Cell*, **166**, 822–839.
- Massagué, J. (2012) TGF β signalling in context. *Nat. Rev. Mol. Cell Biol.*, **13**, 616–630.
- Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T. and Reuter, G. (2004) Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes Dev.*, **18**, 2973–2983.
- Tiwari, V.K., Stadler, M.B., Wirbelauer, C., Paro, R., Schübeler, D. and Beisel, C. (2011) A chromatin-modifying function of JNK during stem cell differentiation. *Nat. Genet.*, **44**, 94–100.
- Flavahan, W.A., Gaskell, E. and Bernstein, B.E. (2017) Epigenetic plasticity and the hallmarks of cancer. *Science*, **357**, eaal2380.
- Epping, M.T., Wang, L., Edell, M.J., Carlée, L., Hernandez, M. and Bernards, R. (2005) The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell*, **122**, 835–847.

10. Rousseaux,S., Debernardi,A., Jacquiau,B., Vitte,A.-L., Vesin,A., Nagy-Mignotte,H., Moro-Sibilot,D., Brichon,P.-Y., Lantuejoul,S., Hainaut,P. *et al.* (2013) Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci. Transl. Med.*, **5**, 186ra66.
11. Finn,O.J. (2018) A Believer's overview of cancer immunosurveillance and immunotherapy. *J. Immunol.*, **200**, 385–391.
12. Gjerstorff,M.F., Andersen,M.H. and Ditzel,H.J. (2015) Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget*, **6**, 15772–15787.
13. Gibbs,Z.A. and Whitehurst,A.W. (2018) Emerging contributions of Cancer/Testis antigens to neoplastic behaviors. *Trends Cancer*, **4**, 701–712.
14. Le Pabic,H., Bonnier,D., Wewer,U.M., Coutand,A., Musso,O., Baffet,G., Clément,B. and Thérêt,N. (2003) ADAM12 in human liver cancers: TGF-beta-regulated expression in stellate cells is associated with matrix remodeling. *Hepatology*, **37**, 1056–1066.
15. Le Pabic,H., L'Helgoualc'h,A., Coutant,A., Wewer,U.M., Baffet,G., Clément,B. and Thérêt,N. (2005) Involvement of the serine/threonine p70S6 kinase in TGF-beta1-induced ADAM12 expression in cultured human hepatic stellate cells. *J. Hepatol.*, **43**, 1038–1044.
16. Roy,R., Wewer,U.M., Zurakowski,D., Pories,S.E. and Moses,M.A. (2004) ADAM 12 cleaves extracellular matrix proteins and correlates with cancer status and stage. *J. Biol. Chem.*, **279**, 51323–51330.
17. Peduto,L., Reuter,V.E., Sehara-Fujisawa,A., Shaffer,D.R., Scher,H.I. and Blobel,C.P. (2006) ADAM12 is highly expressed in carcinoma-associated stroma and is required for mouse prostate tumor progression. *Oncogene*, **25**, 5462–5466.
18. Shao,S., Li,Z., Gao,W., Yu,G., Liu,D. and Pan,F. (2014) ADAM-12 as a diagnostic marker for the proliferation, migration and invasion in patients with small cell lung cancer. *PLoS ONE*, **9**, e85936.
19. Duhachek-Muggy,S., Qi,Y., Wise,R., Alyahya,L., Li,H., Hodge,J. and Zolkiewska,A. (2017) Metalloprotease-disintegrin ADAM12 actively promotes the stem cell-like phenotype in claudin-low breast cancer. *Mol. Cancer*, **16**, 32.
20. Yamaguchi,K., Shirakabe,K., Shibuya,H., Irie,K., Oishi,I., Ueno,N., Taniguchi,T., Nishida,E. and Matsumoto,K. (1995) Identification of a member of the MAPKKK family as a potential mediator of TGF-beta signal transduction. *Science*, **270**, 2008–2011.
21. Solomon,E., Li,H., Duhachek Muggy,S., Syta,E. and Zolkiewska,A. (2010) The role of SnoN in transforming growth factor beta1-induced expression of metalloprotease-disintegrin ADAM12. *J. Biol. Chem.*, **285**, 21969–21977.
22. Ramdas,V., McBride,M., Denby,L. and Baker,A.H. (2013) Canonical transforming growth factor-beta signaling regulates disintegrin metalloprotease expression in experimental renal fibrosis via miR-29. *Am. J. Pathol.*, **183**, 1885–1896.
23. Kim,Y.M., Kim,J., Heo,S.C., Shin,S.H., Do,E.K., Suh,D.-S., Kim,K.-H., Yoon,M.-S., Lee,T.G. and Kim,J.H. (2012) Proteomic identification of ADAM12 as a regulator for TGF-beta1-induced differentiation of human mesenchymal stem cells to smooth muscle cells. *PLoS ONE*, **7**, e40820.
24. Ray,A., Dhar,S. and Ray,B.K. (2010) Transforming growth factor-beta1-mediated activation of NF-kappaB contributes to enhanced ADAM-12 expression in mammary carcinoma cells. *Mol. Cancer Res.*, **8**, 1261–1270.
25. Ruff,M., Leyme,A., Le Cann,F., Bonnier,D., Le Seyec,J., Chesnel,F., Fattet,L., Rimokh,R., Baffet,G. and Thérêt,N. (2015) The disintegrin and metalloprotease ADAM12 is associated with TGF-beta-Induced epithelial to mesenchymal transition. *PLoS ONE*, **10**, e0139179.
26. Ferrand,M., Kirsh,O., Griveau,A., Vindrieux,D., Martin,N., Defossez,P.-A. and Bernard,D. (2015) Screening of a kinase library reveals novel pro-senescence kinases and their common NF-kappaB-dependent transcriptional program. *Aging (Albany, NY)*, **7**, 986–1003.
27. Ma,X., Warnier,M., Raynard,C., Ferrand,M., Kirsh,O., Defossez,P.-A., Martin,N. and Bernard,D. (2018) The nuclear receptor RXRA controls cellular senescence by regulating calcium signaling. *Aging Cell*, **17**, e12831.
28. Glazer,C.A., Smith,I.M., Ochs,M.F., Begum,S., Westra,W., Chang,S.S., Sun,W., Bhan,S., Khan,Z., Ahrendt,S. *et al.* (2009) Integrative discovery of epigenetically derepressed cancer testis antigens in NSCLC. *PLoS ONE*, **4**, e8189.
29. Miotto,B., Marchal,C., Adelmant,G., Guinot,N., Xie,P., Marto,J.A., Zhang,L. and Defossez,P.-A. (2018) Stabilization of the methyl-CpG binding protein ZBTB38 by the deubiquitinase USP9X limits the occurrence and toxicity of oxidative stress in human cells. *Nucleic Acids Res.*, **46**, 4392–4404.
30. Laget,S., Miotto,B., Chin,H.G., Estève,P.-O., Roberts,R.J., Pradhan,S. and Defossez,P.-A. (2014) MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics*, **9**, 546–556.
31. Filion,G.J.P., Zhenilo,S., Salozhin,S., Yamada,D., Prokhortchouk,E. and Defossez,P.-A. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.*, **26**, 169–181.
32. Dulauroy,S., Di Carlo,S.E., Langa,F., Eberl,G. and Peduto,L. (2012) Lineage tracing and genetic ablation of ADAM12(+) perivascular cells identify a major source of profibrotic cells during acute tissue injury. *Nat. Med.*, **18**, 1262–1270.
33. Roussel-Gervais,A., Naciri,I., Kirsh,O., Kasprzyk,L., Velasco,G., Grillo,G., Dubus,P. and Defossez,P.-A. (2017) Loss of the Methyl-CpG-Binding protein ZBTB4 alters mitotic checkpoint, increases aneuploidy, and promotes tumorigenesis. *Cancer Res.*, **77**, 62–73.
34. Ferry,L., Fournier,A., Tsusaka,T., Adelmant,G., Shimazu,T., Matano,S., Kirsh,O., Amouroux,R., Dohmae,N., Suzuki,T. *et al.* (2017) Methylation of DNA ligase 1 by G9a/GLP recruits UHRF1 to replicating DNA and regulates DNA methylation. *Mol. Cell*, **67**, 550–565.
35. Yamada,D., Pérez-Torrado,R., Filion,G., Caly,M., Jammart,B., Devignot,V., Sasai,N., Ravassard,P., Mallet,J., Sastre-Garau,X. *et al.* (2009) The human protein kinase HIPK2 phosphorylates and downregulates the methyl-binding transcription factor ZBTB4. *Oncogene*, **28**, 2535–2544.
36. Collado-Torres,L., Nellore,A., Kammers,K., Ellis,S.E., Taub,M.A., Hansen,K.D., Jaffe,A.E., Langmead,B. and Leek,J.T. (2017) Reproducible RNA-seq analysis using recount2. *Nat. Biotechnol.*, **35**, 319–321.
37. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
38. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.*, **43**, e47.
39. Boehm,J.S., Zhao,J.J., Yao,J., Kim,S.Y., Firestein,R., Dunn,I.F., Sjöstrom,S.K., Garraway,L.A., Weremowicz,S., Richardson,A.L. *et al.* (2007) Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell*, **129**, 1065–1079.
40. Ninomiya-Tsuji,J., Kajino,T., Ono,K., Ohtomo,T., Matsumoto,M., Shiina,M., Mihara,M., Tsuchiya,M. and Matsumoto,K. (2003) A resorcylic acid lactone, 5Z-7-oxozeaenol, prevents inflammation by inhibiting the catalytic activity of TAK1 MAPK kinase kinase. *J. Biol. Chem.*, **278**, 18485–18490.
41. Michishita,E., McCord,R.A., Berber,E., Kioi,M., Padilla-Nash,H., Damian,M., Cheung,P., Kusumoto,R., Kawahara,T.L.A., Barrett,J.C. *et al.* (2008) SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature*, **452**, 492–496.
42. Helmlinger,D. and Tora,L. (2017) Sharing the SAGA. *Trends Biochem. Sci.*, **42**, 850–861.
43. Fournier,M., Orpinell,M., Grauffel,C., Scheer,E., Garnier,J.-M., Ye,T., Chavant,V., Joint,M., Esashi,F., Dejaegere,A. *et al.* (2016) KAT2A/KAT2B-targeted acetylome reveals a role for PLK4 acetylation in preventing centrosome amplification. *Nat. Commun.*, **7**, 13227.
44. Jurida,L., Soelch,J., Bartkuhn,M., Handschick,K., Müller,H., Newel,D., Weber,A., Dittrich-Breiholz,O., Schneider,H., Bhujju,S. *et al.* (2015) The activation of IL-1-Induced enhancers depends on TAK1 kinase activity and NF-kappaB p65. *Cell Rep.*, **S2211-1247**, 00002-9.
45. Duhachek-Muggy,S., Li,H., Qi,Y. and Zolkiewska,A. (2013) Alternative mRNA splicing generates two distinct ADAM12 prodomain variants. *PLoS ONE*, **8**, e75730.
46. Guo,J., Zhang,J., Zhang,X., Zhang,Z., Wei,X. and Zhou,X. (2014) Constitutive activation of MEK1 promotes Treg cell instability in vivo. *J. Biol. Chem.*, **289**, 35139–35148.

47. Leonardi, M., Perna, E., Tronolone, S., Colecchia, D. and Chiariello, M. (2018) Activated kinase screening identifies the IKBKE oncogene as a positive regulator of autophagy. *Autophagy*, doi:10.1080/15548627.2018.1517855.
48. Ellis, J.M. and Wolfgang, M.J. (2012) A genetically encoded metabolite sensor for malonyl-CoA. *Chem. Biol.*, **19**, 1333–1339.
49. Zhou, H., Dickson, M.E., Kim, M.S., Bassel-Duby, R. and Olson, E.N. (2015) Akt1/protein kinase B enhances transcriptional reprogramming of fibroblasts to functional cardiomyocytes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 11864–11869.
50. Yadav, T., Quivy, J.-P. and Almouzni, G. (2018) Chromatin plasticity: A versatile landscape that underlies cell fate and identity. *Science*, **361**, 1332–1336.
51. Tasselli, L., Zheng, W. and Chua, K.F. (2017) SIRT6: Novel mechanisms and links to aging and disease. *Trends Endocrinol. Metab.*, **28**, 168–185.
52. Kawahara, T.L.A., Rapicavoli, N.A., Wu, A.R., Qu, K., Quake, S.R. and Chang, H.Y. (2011) Dynamic chromatin localization of Sirt6 shapes stress- and aging-related transcriptional networks. *PLoS Genet.*, **7**, e1002153.
53. Dominy, J.E., Lee, Y., Jedrychowski, M.P., Chim, H., Jurczak, M.J., Camporez, J.P., Ruan, H.-B., Feldman, J., Pierce, K., Mostoslavsky, R. et al. (2012) The deacetylase Sirt6 activates the acetyltransferase GCN5 and suppresses hepatic gluconeogenesis. *Mol. Cell*, **48**, 900–913.
54. Ajibade, A.A., Wang, H.Y. and Wang, R.-F. (2013) Cell type-specific function of TAK1 in innate immune signaling. *Trends Immunol.*, **34**, 307–316.
55. Thiefes, A., Wolter, S., Mushinski, J.F., Hoffmann, E., Dittrich-Breiholz, O., Graue, N., Dörrie, A., Schneider, H., Wirth, D., Luckow, B. et al. (2005) Simultaneous blockade of NFκB, JNK, and p38 MAPK by a kinase-inactive mutant of the protein kinase TAK1 sensitizes cells to apoptosis and affects a distinct spectrum of tumor necrosis factor [corrected] target genes. *J. Biol. Chem.*, **280**, 27728–27741.
56. Rzeckowski, K., Beuerlein, K., Müller, H., Dittrich-Breiholz, O., Schneider, H., Kettner-Buhrow, D., Holtmann, H. and Kracht, M. (2011) c-Jun N-terminal kinase phosphorylates DCP1a to control formation of P bodies. *J. Cell Biol.*, **194**, 581–596.
57. Handschick, K., Beuerlein, K., Jurida, L., Bartkuhn, M., Müller, H., Soelch, J., Weber, A., Dittrich-Breiholz, O., Schneider, H., Scharfe, M. et al. (2014) Cyclin-dependent kinase 6 is a chromatin-bound cofactor for NF-κB-dependent gene expression. *Mol. Cell*, **53**, 193–208.
58. Kveiborg, M., Fröhlich, C., Albrechtsen, R., Tischler, V., Dietrich, N., Holck, P., Kronqvist, P., Rank, F., Mercurio, A.M. and Wewer, U.M. (2005) A role for ADAM12 in breast tumor progression and stromal cell apoptosis. *Cancer Res.*, **65**, 4754–4761.
59. Sakurai, H. (2012) Targeting of TAK1 in inflammatory disorders and cancer. *Trends Pharmacol. Sci.*, **33**, 522–530.
60. Singh, A., Sweeney, M.F., Yu, M., Burger, A., Greninger, P., Benes, C., Haber, D.A. and Settleman, J. (2012) TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. *Cell*, **148**, 639–650.
61. Totzke, J., Gurbani, D., Raphemot, R., Hughes, P.F., Bodoor, K., Carlson, D.A., Loisel, D.R., Bera, A.K., Eibschutz, L.S., Perkins, M.M. et al. (2017) Takinib, a selective TAK1 inhibitor, broadens the therapeutic efficacy of TNF-α inhibition for cancer and autoimmune disease. *Cell Chem Biol*, **24**, 1029–1039.
62. Ji, Y.-X., Huang, Z., Yang, X., Wang, X., Zhao, L.-P., Wang, P.-X., Zhang, X.-J., Alves-Bezerra, M., Cai, L., Zhang, P. et al. (2018) The deubiquitinating enzyme cylindromatosis mitigates nonalcoholic steatohepatitis. *Nat. Med.*, **24**, 213–223.
63. Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L. and Devys, D. (2014) The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev.*, **28**, 1999–2012.

Supplemental material for Naciri et al.

**“Genetic screens reveal mechanisms for the transcriptional regulation of
tissue-specific genes in normal cells and tumors”**

Contents :

Supplemental Figure Legends

Supplemental Table Legends

Supplemental Figures S1-S7

Supplemental Tables 1-7

Supplemental Figure Legends

Supplemental Figure 1, Related to Figure 1.

(A) Validation of Nanostring probes with positive controls. The probes for each of the 42 tissue-restricted genes were tested on primary IMR90 cells (black dots), the transformed derivative SW39 cells (green dots), and on a positive control mRNA extracted from 10 mixed human cancer cell lines. The dotted line represents a 2-fold change over control. (B) Verification of kinase expression in the signaling screen: Nanostring counts of the Myr-Flag tag in non-infected cells (IMR90 or its transformed derivative SW39), and in cells infected by each of the 192 activated kinases (yellow dots). (C) Validation of siRNA efficiency on eight selected genes. Nanostring probes were used to measure the abundance of mRNA for each of the 8 indicated genes after transfection of siRNA pools targeting the gene in question (yellow bars), or transfection of non-targeting siRNA pools (green bars). Non-transfected cells are also included (black bars).

Supplemental Figure 2, Related to Figure 2.

(A) Relative expression of ADAM12 and TAK1 in MRC5 primary lung fibroblasts after infection with viral vectors expressing either wild-type TAK1 or the catalytic dead TAK1 (TAK1 CD). The statistical analysis was performed with a one way ANOVA followed by a Dunnett's test except for mice experiment where Mann Whitney was performed (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). (B) RT-qPCR representing the relative expression of TGF- β target genes in the presence or absence of TAK1 kinase inhibitor 5Z. MRC5 cells were pre-treated with 0.3 μ M 5Z or DMSO for two hours, followed by stimulation with 5ng/mL TGF- β for six hours. (C) Western blot showing the phosphorylation of SMAD3 in MRC5 even under 5Z treatment. Tubulin (TUB) served as a loading control. (D) Western blotting on ADAM12 in whole cell extracts of tibialis muscles after the indicated treatments. CD71 is a loading control that underwent Concanavalin A enrichment, like ADAM12. MW: lane with the molecular weights. (E) Quantification of panel D was performed using Image Studio lite (Licor) with CD71 as a loading control. All experiments have been performed three times while mice experiments were conducted twice.

Supplemental Figure 3, Related to Figure 3.

(A) RT-qPCR showing relative expression of ADAM12, KAT2A and SIRT6 in MRC5 cells transfected with siCtl (non-targeting siRNA), siKAT2A and/or siSIRT6 as indicated. These siRNAs are independent from the ones used in the screen. The statistical analysis was

performed employing a two-way two tailed ANOVA test followed by Dunnett's test. **(B)** Western blot analysis showing down regulation of KAT2A and SIRT6 after siRNA transfection in MRC5, Tubulin (TUB) was used as a loading control. **(C)** Western blot showing induction of the ADAM12 protein after siRNA against the indicated factors. All experiments were performed three times or more.

Supplemental Figure 4, Related to Figure 4.

Primary lung fibroblast IMR90 **(A)** and MRC5 **(B)** were transfected with siCtrl or siKAT2A and treated every day with 0.3 μ M 5Z or DMSO for 3 days. RT-qPCR shows relative expression of ADAM12 and KAT2A. The statistical analysis was performed with a one way ANOVA test followed by a Dunnett's test. All experiments were performed three times or more.

Supplemental Figure 5, Related to Figure 5.

(A) RT-qPCR showing relative expression of ADAM12 after treatment of BT549 cells with the TAK1 inhibitor 5Z at 0.3 μ M concentration for 4 days. The statistical analysis was performed with a one way two ANOVA test followed by a Dunnett's test. **(B)** Western blot showing ADAM12 expression after 5Z treatment in BT549. For ADAM12, concanavalin A enrichment was performed. **(C)** Co-immunoprecipitation of endogenous TAK1 from SUM159PT after cellular fractionation revealed interaction with KAT2A in the cytoplasm and the nucleus. TAB1, co-factor of TAK1 served as a positive control. Histone H3 was used as a marker of the nuclear compartment, and tubulin as a marker of the cytoplasmic compartment.

Supplemental Figure 6, Related to Figure 6.

Co-immunoprecipitation of endogenous TAK1 from MDA-MB-231 cell line reveals interaction with KAT2A.

Supplemental Figure 7, Related to Figure 7.

(A) Random sampling procedure to test the validity of the TAK1 signature. **(B)** ADAM12 expression in lung tumors and normal lung from the TCGA database. The boxplots contain 50% of the values, with a notch at the median value, and a diamond at the average value. The whiskers depict the first and last quartiles, and outliers appear as black dots. **(C)** ADAM12, KAT2A, TAK1 expression in normal lung tissue and a subset of lung tumors, selected based

on their low (1st decile, green) or high (9th decile, red) ADAM12 expression. **(D)** Unsupervised clustering of lung tumors according to the TAK1 signature almost perfectly segregates tumors according to ADAM12 expression. **(E)** ADAM12 expression in colon cancer samples and normal colon tissue samples from the TCGA database. **(F)** ADAM12, KAT2A, TAK1 expression in normal colon tissue and a subset of colon tumors, selected based on their low (1st decile, green) or high (9th decile, red) ADAM12 expression. **(G)** Unsupervised clustering of colon tumors according to the TAK1 signature almost perfectly segregates tumors according to ADAM12 expression. **(H)** TAK1-activation score versus ADAM12 or KAT2A mRNA expression in lung cancer samples from TCGA database. The statistical analysis was performed employing a one way ANOVA test, following by a Tukey HSD test (***) denotes $p < 0.001$. **(I)** As in panel H, but on colon tumors.

Supplemental Table Legends

Table S1: List of 192 activated kinases used in the signaling screen.

Table S2: List and target sequence of 72 Nanostring probes.

Table S3: List and characteristics of 42 tissue-restricted genes analyzed in the screens.

Table S4: List of 160 chromatin regulators targeted in the siRNA screen.

Table S5: Sequence of qRT-PCR primers used.

Table S6: Sequence of siRNAs used.

Table S7: List of 190 genes constituting the TAK1 signature.

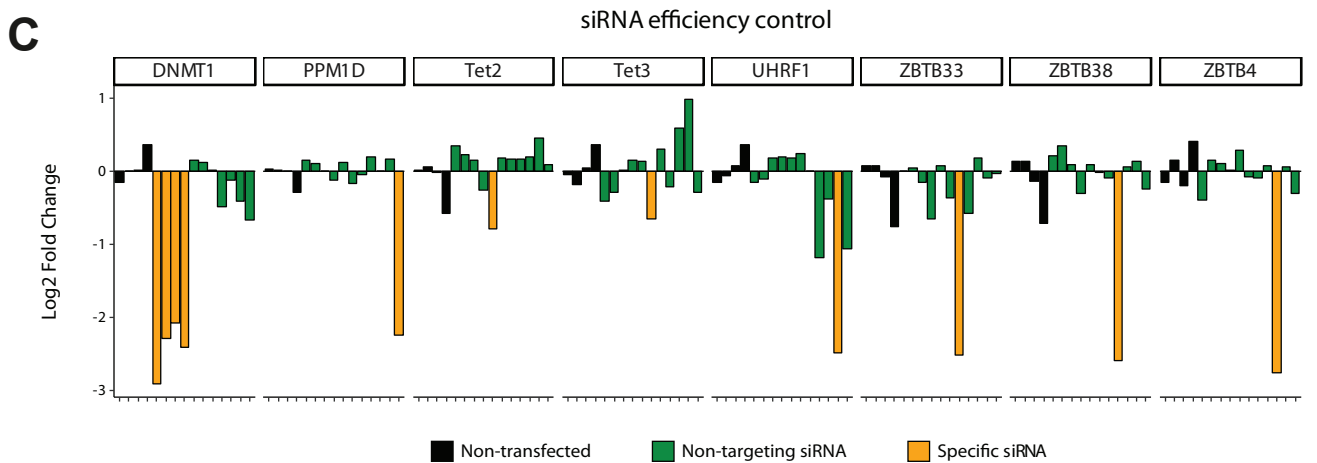
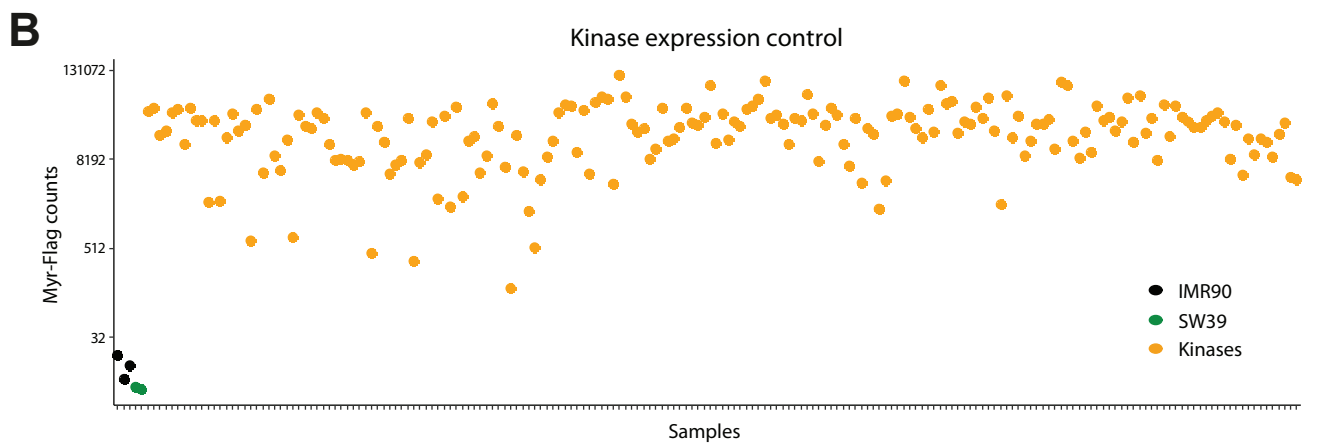
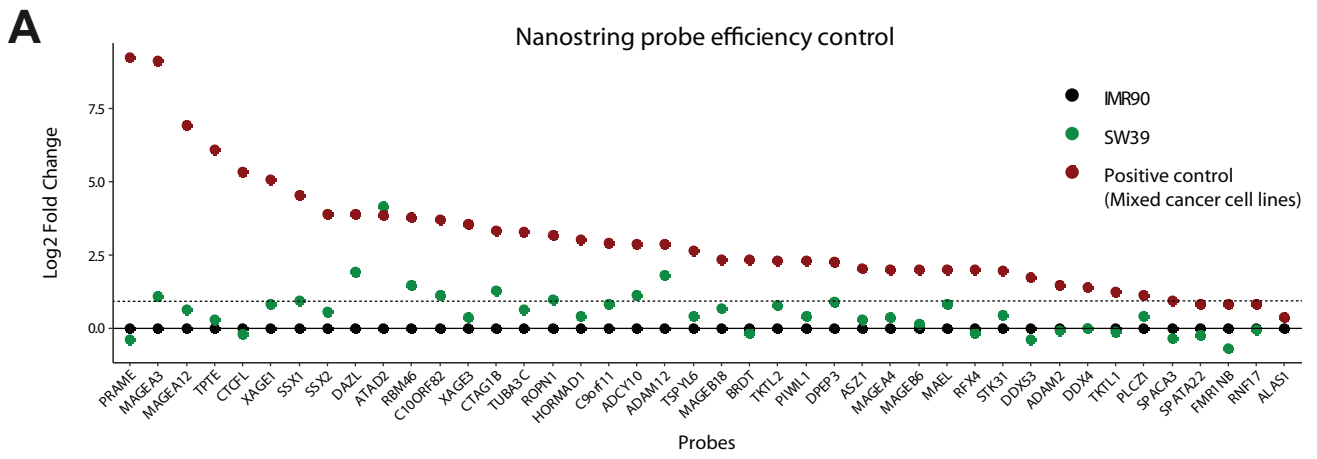


Figure S1
Naciri et al.

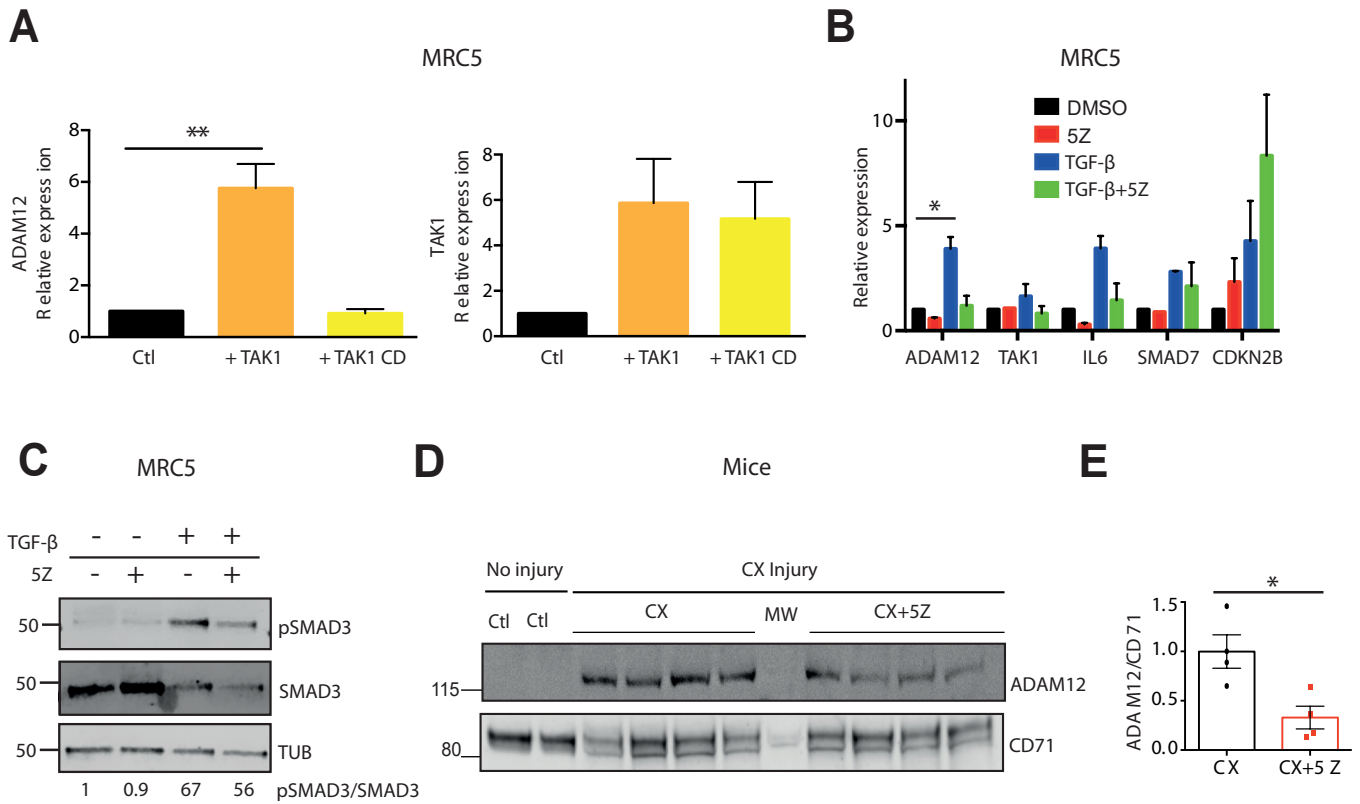


Figure S2
Naciri et al.

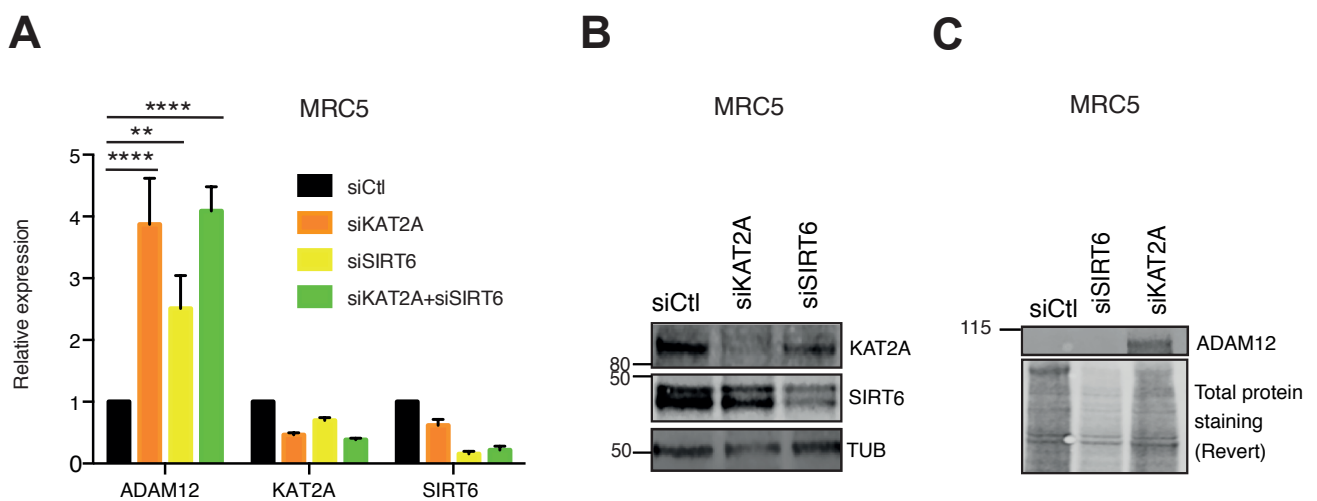


Figure S3
Naciri et al.

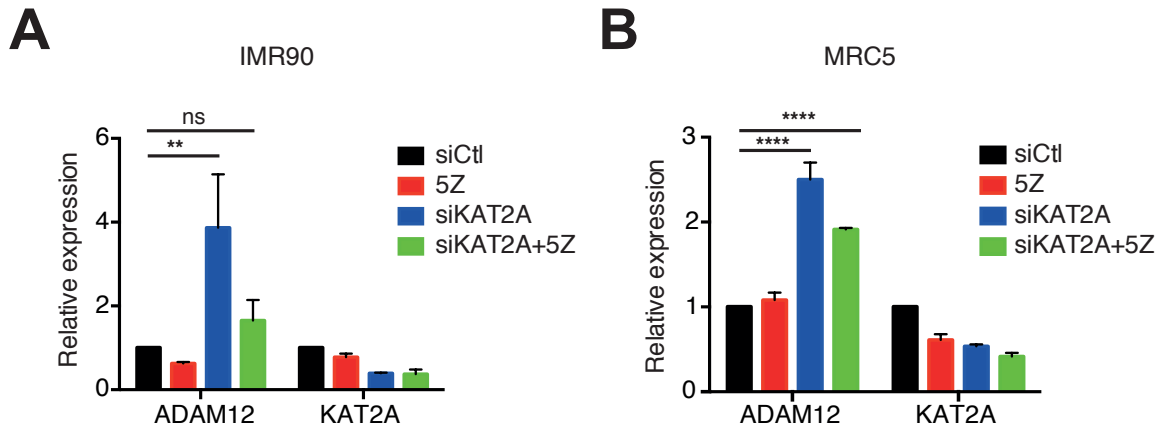


Figure S4
Naciri et al.

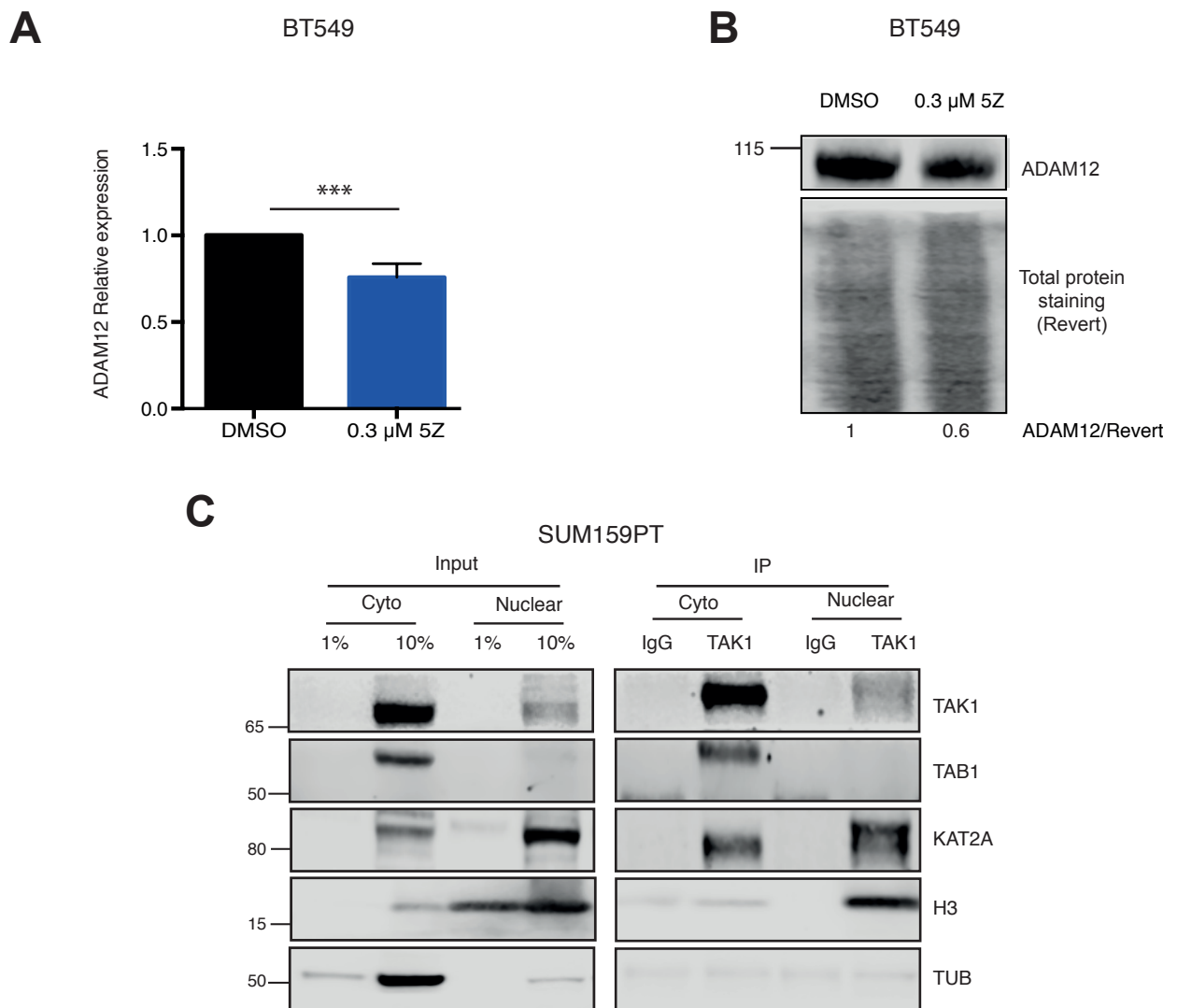


Figure S5
Naciri et al.

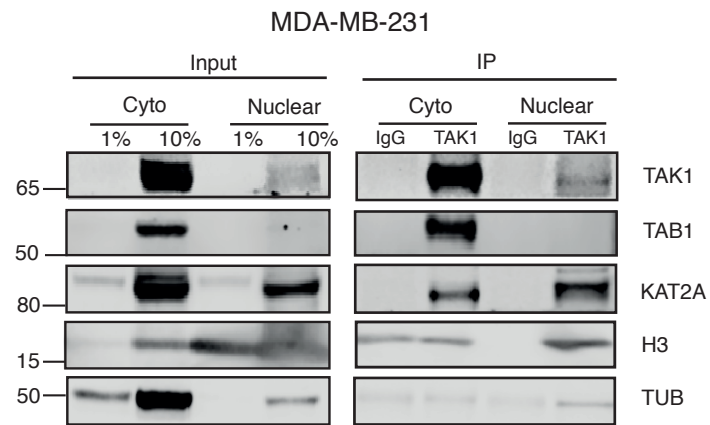


Figure S6
Naciri et al.

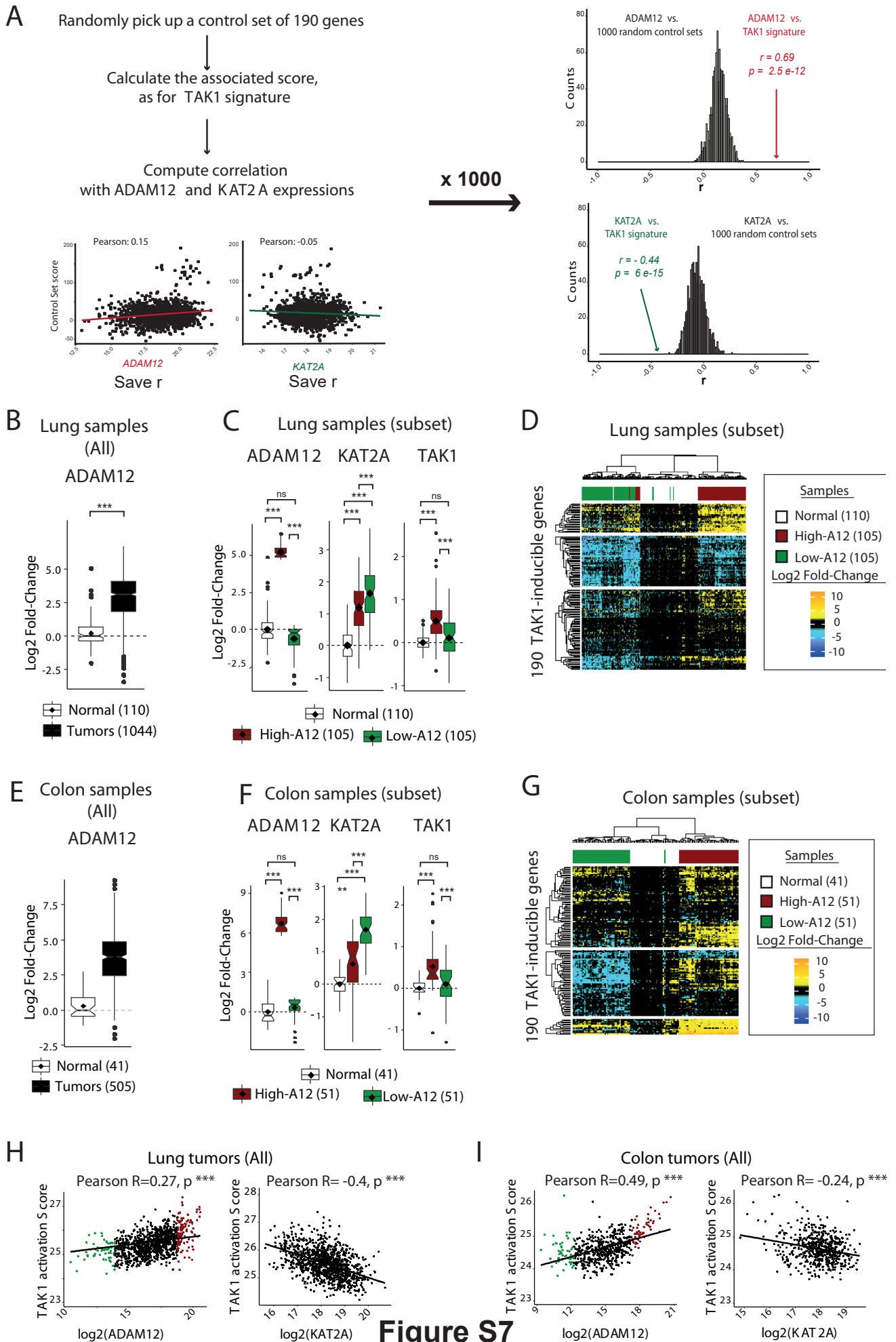


Figure S7
Naciri et al.

Table S1 : List of the kinases used in the screen

192 kinases genetically activated by addition of Myristoylation sequence

All in pBabe vector (Hahn/Zhao; Addgene#1000000012)

AAK1	CKMT1A	HK1	MVK	PIP5K1B	RPS6KL1
ACVR1	CKMT2	HK2	NADK	PIP5K2A	RPSK6A3
ADCK4	CKS1B	HK3	NEK11	PIP5K3	SGK
ADCK5	CKS2	IHPK2	NEK3	PKM2	SNF1LK
ADPGK	CLK1	IKBKE	NEK6	PKN1	SPHK2
ADRBK1	CLK2	ILK	NME7	PKN2	SRPK2
ADRBK2	CLK3	ITK	NTRK3	PLAU	STK17B
AKT1	CMPK	ITPK1	NUAK2	PLK1	STK3
AKT3	CSNK1A1L	ITPKB	OXSRI	PLK2	STK32A
AMHR2	CSNK1E	LCK	PACSIN1	PLK3	STK32B
AURKA	CSNK1G1	LIMK1	PAK4	PLK4	STK32C
AXL	CSNK1G2	LIMK2	PAPSS1	PMVK	STK33
BLK	DAK	MAP2K5	PBK	PNKP	STK38L
BMX	DGKG	MAP2K6	PCK2	PRKAA1	STK4
BTK	DGUOK	MAP2K7	PCTK1	PRKACB	STK40
CALM2	DLG5	MAP3K14	PCTK2	PRKACG	SYK
CAMK1G	DYRK2	MAP3K6	PCTK3	PRKAG2	TAOK3
CAMK2B	DYRK4	MAP3K7	PDIK1L	PRKAR2A	TBK1
CAMK2D	EPHA4	MAP3K8	PDK1	PRKCD	TEC
CAMK4	FASTK	MAPK12	PDPK1	PRKCI	TESK1
CAMKK1	FGFR1	MAPK13	PDXK	PRKCZ	TIE1
CAMKV	FGR	MAPK14	PFKL	PRKRA	TK1
CDC2	FRK	MAPK6	PFKM	PTK2	TNK2
CDK2	GAK	MAPK7	PI4K2B	RET	TSSK1B
CDK4	GALK2	MAPKAP1	PIK3CB	RIOK1	TSSK6
CDK5	GCK	MAST1	PIK3CG	RIOK2	TTK
CDK7	GK	MATK	PIK3R3	RIOK3	TYK2
CDK9	GK2	MELK	PIK3R5	RPS6KA2	UCK2
CERK	GRK5	MKNK1	PIK4CA	RPS6KA5	ULK4
CHEK1	GRK6	MOBK1A	PIK4CB	RPS6KA6	VRK2
CKB	HCK	MOBK2A	PIM1	RPS6KB1	VRK3
CKM	HIPK1	MPP1	PIP5K1A	RPS6KB2	YES1

Table S2 : List of Nanostring probes

Probe number	symbol	Probe class	Accession #	Target Sequence
1	ADAM12	Tissue-restricted	NM_003474.4	GTAAGTATGTCCTCCGCTCGAAATACACGGTAATTCTGGTCTACTTACTACCATGGACATGTACGGGATATTCTGATTGATCAGCAGTCAGTCTCAG
2	ADAM2	Tissue-restricted	NM_001464.3	AGGATTGTGCCCTTATTGGAGAAACATGCTGTGATATTGCCATGTAGATTTAAAGCCGGTTCAAACCTGTGCTGAAGGACCATGCTCGAAAACTGCT
3	ADCY10	Tissue-restricted	NM_001167749.1	CAAAACTGTCTTTATGGTGCAGCTATAGACAGACATGGCGCTTCTGCTGTCTCAGAAATATGTCACCCGGCCTCTCTCTGCTCTGCGAGTCTTA
4	ALAS1	Tissue-restricted	NM_000688.4	GGGGATCGGGATGGAGTCAATGCAAAAAATGGACATCATTTCTGGAACACTTGGCAAAGCCTTTGGTTGGTGGAGGGTACATCGCCAGCAGGATTC
5	ASZ1	Tissue-restricted	NM_130768.2	GGTTCCGGTCTTTGGACAGAGGTGCTAATGCAAGCTTTGAAAGGATAAGCAAAGTATTTTGTAACTCATGTTCTCTCATGGTCCAGAGGAAAC
6	ATAD2	Tissue-restricted	NM_014109.3	GAAAAACCTCGTACCAGAGAAGCCCAACATATTTATAGTGGCCAGCTTCTCTGCAAGACCAAGATACCGATTATCTCCGACGAGCAAGAAAGTC
7	BRDT	Tissue-restricted	NM_001726.3	ACAGATAGGATATTGTGCAAGACACAACCTCTGCAACTACTCCCTTGTTCATCAGACCACACTTCAATGTAATGCCAAATACACCAACTTA
8	C10ORF82	Tissue-restricted	NM_144661.2	AAGATGTATCTAGAGCCACTGCTCCGCAAAAGTATGCAAGAGGCTGAGAAGCGCAGAGTCTCCAAAGGAGGGAACCTTAAAGTGGGCTTCCAAAACCTG
9	C9orf11	Tissue-restricted	NM_001161585.1	CCAGAGCTGCCACGATGTTACTTCTTCCATCCTGAAAGGTTTTCAGATACATCTTCCAAAGAGTGCAGAGCAGCAGCATTCTTTGGTACCATT
10	CTAG1B	Tissue-restricted	NM_001327.2	GCGGGGCCAGGGGCGGCGCTGCTGAGTTCTACTCTGCTGCTTCTCGCGACCCCTATGTTGAGGAGAGCTGACCCGCGCCGACGAGCTGGCCCA
11	CTCFL	Tissue-restricted	NM_001269042.1	TTAACACCCACACAGGAACCCAGGCGCTCAAGTGAACGACTGCAACATGGCATTTGTCCAGTGGAGAACTCGTCCGACACAGCGCTATAAACATAC
12	DAZL	Tissue-restricted	NM_001351.2	ACCTCCGGCTTATTCACTGTTAACTACCAGTAAATGAAGTGTCCAGGAGCTGAAGTTGTGCCAAATGAATGTCCAGTTCATGAAGCTACTACCC
13	DDX4	Tissue-restricted	NM_024415.2	AGAAAGTAGTGATACTCAAGGCCAAAGATGACCTACATACCCCTCTCCACTGAGGATGAGGAGCTTCTTGTGACCAATCAGACAGGCAATAAC
14	DDX53	Tissue-restricted	NM_182699.2	ACAAACCCGAAACAGGAAAACTTGTCTTCTAATGCTGGGTTTATTCTCTGATTCTCAACCAATATCTAGAGAGCAAGGAATGGGCTGGGATG
15	DPEP3	Tissue-restricted	NM_022357.3	CAGCGGATGACAAGCTTTGGTGAAGAAAGTAGAGAGGTTGAACCGCCTGGCATGATGATGATTGCTTCTGATCGGACACTTGTAAAGAAAGG
16	FMR1NB	Tissue-restricted	NM_152578.2	TCCTGGTATGCTGCCATTATTGCGCTCTCTTTCTGGAGGCGAACCAGGCGGATGATTTCAAAGGCGAGACAAACAGAGTTGTAACGGTGGTAA
17	HORMAD1	Tissue-restricted	NM_032132.3	ATTCAAATACACCAATAATGGACCTCATGGAATTTAATGAAAAACCAAGCAAGAACTAGCATGTTGCTACTGACCAAGAAAGCAAGCAAT
18	MAEL	Tissue-restricted	NM_001286377.1	CATCAGAAATCAGGCAAGATCTCAACTTCTCACTGTAGAGGACCTTGTAGTGGGGATCTACCAAAAAATTTCTCAAGGAGCCCTCTAAGACTGGAT
19	MAGEA12	Tissue-restricted	NM_001166386.1	CCCACTACCATCACTATCTCTGGAGTCAATCCGATGAGGGCTCCAGCAACGAAAGCAAGGCGCAACACTTCTGCTGACCTGGAGGAGCT
20	MAGEA3	Tissue-restricted	NM_005362.3	ACTGTGCCCTGAGGAGAAAACTGGGAGGAGCTGAGTGTGTAGAGGTTTGAAGGGAGGGGAAAGACAGATCTGGGGGATCCCAAGAACTGCTCAC
21	MAGEA4	Tissue-restricted	NM_001011548.1	TGATGGCTCTGGTAAATCAGATCTTTCCAAAGACAGGCTTGTGATATCTGCTGGCACAATTGCAATGGAGGGCGACAGGCTCTGAGGAG
22	MAGEB18	Tissue-restricted	NM_173699.3	AGAAATAAATCTCAGGATCCGGCAGGTGCAAAAGGTCCTCTCATCAATCAACTGCTGCACTATTGTCTGTAGTCCCTGACAAGATAAATATGCC
23	MAGEB6	Tissue-restricted	NM_173523.2	TGGGAGTCTCTGGGCTGTGGGGATATGATGGGATCTGCTCAATCAATCTATGGGATGCTCGGAAGTCACTACTGAAGATTTGGTCAAGATAAGT
24	PIWIL1	Tissue-restricted	NM_001190971.1	TAACCTCAGAGCAAGGACAGCTGAAGTGGGACGACACTTATTGATTCATAAAAAACGATAAATGTTCAAAGGAGCTTCGAGACTGGGTTGAGCT
25	PLCZ1	Tissue-restricted	NM_033123.3	TGCATTCATGACATCTGACTACCCAGTGGTCTCTTTAGAAAATCACTGCTCCACTGCCCAACAAGAAAGTATGGCAGACAAATGCGAGGCTACTTT
26	PRAME	Tissue-restricted	NM_006115.3	AGGACCTGGTCTTGTAGTGTGGGATCAAGGATGATCAGTCTGCTCCCTCTGCTTCCCTGAGCCACTGCTCCCACTCAACAACCTTAAGCTTCTA
27	RBM46	Tissue-restricted	NM_001271717.1	CTTATCCAGGCTATCTTGTCCACCAAAATCACTGCTAATGGCAGCCATGTTGACAGCGGCTATGATCTCCAATCAGGCTCCTCTCTCTGAAG
28	RFX4	Tissue-restricted	NM_213594.1	CTGCCAGCCAGCTGCTGAGGAGAAGTTCCTACCTTATTATGATGTACAGAACACACTGTGAGAAATCTGGACACTGTAATAAGGCAACTTTG
29	RNF17	Tissue-restricted	NM_001184993.1	CCAGCACCAGACAGATAGTGACATATATGAGGATGAAACAGCATCACTTATATGCTTGGTGAAGTGGGGCTTGGCAGATAAAGATGAATAAGTGC
30	ROPN1	Tissue-restricted	NM_017578.2	TCCTGCCATCTACATCTGCTGCAAACTGCAAAAGTGTGCTGCAAAAGAGGCTGGTGAACCTCTGCTGAGGATGATCTGATGATGATGATGATGAT
31	SPACA3	Tissue-restricted	NM_173847.3	TGCACTCAAGCCCTGTTCTCTCTCTGTGAGTGGACACGAGGCTGTTGAGTCTGCTGCTATCCAAAGCTCAGCTCTGAGCCAGAGTGTGTGGT
32	SPATA22	Tissue-restricted	NM_001170695.1	ACAGATGGGCTGGGAAGCTGTAATCCAGAGTGGCTGCTGTAATGAAAAACAGTGGACCCGGGCAAAATACCACTAGTCTTCTGCTCTGCTGAGAA
33	SSX1	Tissue-restricted	NM_005635.2	CACITTTGCACCAACTGCGCAACTGCTGCTCCGCTGCAACACTGCTTGTGTTGAGTGAAGCTCTCTGCTGGCCATGAAACGGA
34	SSX2	Tissue-restricted	NM_003147.4	TACCTTAGAAAAAAGTATGCTGGTACTCTGATGGAACAGCATACCTCTCTCCCAAGTGACTACTAGGCGAGTCTGAGTGTGATTAAT
35	STK31	Tissue-restricted	NM_031414.2	GGATTTGACACAGAAAGTGAAGTGTAGAGAGGAGCAGCTACCATAGAGCTGGAGAGAAGCTGAAGGAGACTGAGGCTGCTTCTGATTAATC
36	TKTL1	Tissue-restricted	NM_001145933.1	ACGTTTGGAGGCTGTCAGGACGACGAAAGCTTGTAGGATGTCCTGTGCTGTTGATGAGAGCTCCACTGTACTGTTCAAGTCAATGTTAAT
37	TKTL2	Tissue-restricted	NM_032136.4	TGGTTTGGCTCTGCTAACTGGGCGTCAAAATGAAGAGTATTGTTCTGTAGGTTGACAGGATGAATCCACTTTTCTGAGATATTCAGGAAGAA
38	TPTE	Tissue-restricted	NM_199259.2	TAGCTCCGCGCCGAGAGAAATGTTGACACGACGACCAAGACTCAGACTGTGTTATTCTAGCAGCTGAACACCCAGGCTCTTCTGACCGGAGTG
39	TSPYL6	Tissue-restricted	NM_001003937.2	TGGGTCACATCTCTAAGCGAAGCGAATGATGCTGAGTGTACTGATTTACTTGTGTTGGACTCGTGTGTTCCAGGACCTCTGCTGCTGGCTCC
40	TUBA3C	Tissue-restricted	NM_006001.2	GGGCGTGCAGCTGCGAGCGGGGTTGAGTCAAGTAGTAGCGTGGGCTGGCAGCGGAGGAGCTCAACATGCGTGAAGTATCTATCCAGTGGG
41	XAGE1	Tissue-restricted	NM_001097592.2	TGATCTGCAAGAGCTGATCAGTCAAAACCCGGGATAAATCTGAAATTTGGTTCCGCGTCAAGGTGAAGATAAATACCTAAAGAGCAACTGTAAAA
42	XAGE3	Tissue-restricted	NM_133179.2	GGTGTAGGGTTCGCTTCTGCTGTGACTTTTTCTGCTCCACTGAGACGAGCTGTGGAATATGATTGGCGAGGAGATCAACATATAGGCCTA
43	DNMT1	Control	NM_001379.2	CAAAACCAATCTATGATGATGCCACTCTTGAAGTGGTGTAAATGGCAAAATCTTGGCCCATAAATGAATGGTGGATCACTGGCTTGTATGGAGG
44	PPM1D	Control	NM_003620.2	CCCACTCTGACCTCAGAAGCAAGTATATTTGGGGAGTGAATGCTTGGAAATGATTCACCAACAGAGTCCGATCTCAATGTCCAGGAC
45	Tet1	Control	NM_030625.2	AATGCCAATCAGAAAGCCATCTTGTACCCAGCCCTCTCCACTAACCAGTGTGCTAACGTGATGGCAGGCGATGACCAATAACGGTTTCAGCAGG
46	Tet2	Control	NM_001127208.2	CTCATAATGCAAAATGGGACTGAGGAAAGTACAGAATAAATCGTAGAAATCCCTTATGTCAGACCATGAAATCAAGTGCATGCAAAATACAGGT
47	Tet3	Control	NM_144993.1	ATCGCTTCCAGCAGGTTCTCAAAATAGTGTGCTGACTCAAGAAATCATCTGACCATGACCAAGTCTGCTGAGTAAAGGTTGCTGATGTTGG
48	UHRF1	Control	NM_001048201.1	TGTGGAGCTGTGATGTTCTGCTGTCAAGTCTCAGAACTGCTGCAACACTCTGCAAGCTTAAAGAGGCGACAGGATCAGTCTCTCTGCGGTTCTGGCC
49	ZBTB33	Control	NM_006777.3	AGTCAGTACTGGGTTGTGTAATCTGCTAGTCCAGCGCAACAGCAACTCTGCTACCTCCCTCTGAAATAGCCATGCGACAGCTCTCTACTG
50	ZBTB38	Control	NM_001080412.2	TTTGCCAGACAGGACCATGTTAAATTTGTAATGGCAAATGCTCTACAGTTCGCTGTTGTTGCAAACTGATTTATGACCTTATAGCTCCCGA
51	ZBTB4	Control	NM_020899.3	ACCTAATGACCCGCTTGTCTGAAGCTTCTCTAAGCCCTTCCAGTGTCTCTAGCACATCCATCTTGTGGCCAGGCGCTGACAGCAGCCATT
52	MyrFlag	Normalizer	NA	ATGGGGTCTTCAAATCTAAACGAAAGCCACAGCCAGCGCGGCGAGGATCCGAGGTTACTTGACTACAAGACGATGACGACAAGCAATGAC
53	NeoR	Normalizer	NA	ACGCAGGTTCTCCGGCCGCTGGGTTGAGAGGCTATTCGGCTAGACTGGGCACAACAGCAATCGGCTGCTGATGCGCCGGTTCGGGCTGTGAGC
54	PuroR	Normalizer	NA	GCCACGCGCACACCGTTCGATCCGACCCGACATCGAGCGGCTCACCGAGCTCAAGAACTCTCTCAGCGCGCTGGGCTGACATCGGCAAGGTTG
55	B2M	Normalizer	NM_004048.2	CGGGCATTCTGAAGCTGACAGCATTCGGGCGCAGATGCTCGTCCGTTAGCTGTGCTCGGCTACTCTCTTCTGCTGCGGACTACTCA
56	PGK1	Normalizer	NM_000291.3	ATTGTCAAAGACTAATGTCAAAGCTGAGAAGAAAGTGTGTAAGATTACCTTGGCTGTTGACTTGTCACTGCTGACAAGTTTGTAGAGAAATGCCAAGA
57	TBP	Normalizer	NM_001172085.1	ACAGTGAATCTTGGTTGAACTTGAACCTAAGACCATGCACTTGTGCCCCAAGCCGGAATAAATCCCAAGCGGTTTGTCTGCGGTAATCATGAGGA
58	TUBB2A	Normalizer	NM_001069.2	AGGACGAGGCTTAAAACTCTCAGATCAATCGTGCATCTGTGAACTCTGTTGCTCAAGCATGGTCTTCTACTTGTAAACTATGCTGCTCAGT
59	NEG_A	internal controls	ERCC_00096.1	AACCCGCATACGGCCGATTTGCGCAGCCGGGTCGATTAATAACACCGTGAATCTCAGCTAACCCGACGAGTTTGTCTCTGGATTCTGAGCCCG
60	NEG_B	internal controls	ERCC_00041.1	GCACTGGCATTGGTCTTTCAGGAGCCATACAGAAGCTGTTTATATGAAGAAACATGGATCATTGGAAGTCAATGGGGAAACCTTGTATGATGCGGCG
61	NEG_C	internal controls	ERCC_00019.1	GTACAGGCTGCTGGCTATGTTTCTCTCAAGCTGACTTGCAGGATAGAGGTCGGTTCGATCTAATTCGGAGATAAATATCCAGCAAGCACTC
62	NEG_D	internal controls	ERCC_00076.1	AGAGATCAGCTGGACCAAGCTGATTGATTACCGGACTGGCCGTAAGTGTGCTGCCCGAGTAGATCTCTAGATCCGGCTCAAAATTCCTGCGGTGCTCT
63	NEG_E	internal controls	ERCC_00098.1	CCAGATGACCTTCTCCATAACTCTAATCTGAGCAGGAGGCTGATTAATTTCCGCTCCACTGCAACCCGACCGTGTGAACGACGCGCAACT
64	NEG_F	internal controls	ERCC_00126.1	GGGCTTACCGGCTGTAAGCTCACTCAACTCCAGTACAGAGTGGTTCGTTAACCCGCAATGAGAGGCGCTACACCCGTCAGAATTAAAGCTATGGGG
65	NEG_G	internal controls	ERCC_00144.1	ACCCGATGAACTTGGCCGCTGGGAAATGTTAAGGCTCTGGCAGCCTTATCATTGCGAGCTTCTGTCAGCGGCTGAGGATGATGATGATGATG
66	NEG_H	internal controls	ERCC_00154.1	TTGGTCCGAGGAGCTATAGAAACGATGGCAGCGCTATTACAGCTTATTGGTATGGAGTAAGAGCGGAAACTGGGCTGATGATGATGATGATG
67	POS_A	internal controls	ERCC_00117.1	TCAGGCTTCCCTTACTAATGGCGGCTTGAACGGCCTTGGGGAATGTCACTATTGAGGACCCGTTGACCCCTCAGAGATATACCATCCGCTAT
68	POS_B	internal controls	ERCC_00112.1	ATGAAAGCGCTGACTTATGATAAGACTCAGTACAGTCTCGCCGATTTGATTTGGGAGCTGCGCCATGACGCGCACTACTTGAAGCAATGCT
69	POS_C	internal controls	ERCC_00002.1	GCCCTTTCGCTGGCTCTGGGGTTATAGCTTTTCACTGCTGACCGGCTAGCACACACTCTGTTGACTAGGCGCATGCGCATCAGATGATGCT
70	POS_D	internal controls	ERCC_00092.1	TACCTGGATTTGGCGATCTTGGTAAAGCGGAAAGACTCGGAGGCGCCGCTATTTGCGATCTTCCATGTCGATGCGGCTGCTGATGACT
71	POS_E	internal controls	ERCC_00035.1	GGTTGAATTTGAGCGGATGGGCTCAACTGCTCTGTAACCGGTAGATACAGGGCATACGAGCTCCCTATTTAAACGGCATCCCGGCTAGTGTCCGTC
72	POS_F	internal controls	ERCC_00034.1	ACGAAGCTGTGCGCGCAGCAAGTACCTGCTTGAAGAGCGAATTAACCCACGACCGTGTTCACCCCTGGCCGCTCTCAACCTGATGATCA

Table S3 : List of tissue-restricted genes analyzed

	Gene	Tissue with highest expression	Cancer/Testis Antigen?	CpG island?	CpG island methylated in IMR90?
1	ADAM12	Placenta	No	Yes	Partially
2	ADAM2	Testis	Yes	No	
3	ADCY10	Testis	No	No	
4	ALAS1	Adrenal Gland/Liver	No	Yes	No
5	ASZ1	Testis	Yes	Yes	Yes
6	ATAD2	Testis/Bone marrow	Yes	Yes	No
7	BRDT	Testis	Yes	Yes	Yes
8	C10ORF82	Testis	No	Yes	Partially
9	C9orf11	Testis	No	No	
10	CTAG1B	Testis	Yes	No	
11	CTCF	Testis	Yes	Yes	Yes
12	DAZL	Testis	No	Yes	Yes
13	DDX4	Testis	No	Yes	Yes
14	DDX53	Testis	Yes	No	
15	DPEP3	Testis	No	Yes	Yes
16	FMR1NB	Testis	Yes	Yes	Yes
17	HORMAD1	Testis	Yes	No	
18	MAEL	Testis	Yes	Yes	Yes
19	MAGEA12	Testis	Yes	No	
20	MAGEA3	Testis	Yes	Yes	No data
21	MAGEA4	Testis	Yes	No	
22	MAGEB18	Testis	Yes	No	
23	MAGEB6	Testis	Yes	No	
24	PIWIL1	Testis	Yes	Yes	Yes
25	PLCZ1	Testis	No	No	
26	PRAME	Testis	Yes	Yes	Yes
27	RBM46	Testis	Yes	Yes	Yes
28	RFX4	Testis	No	Yes	No
29	RNF17	Testis	No	No	
30	ROPN1	Testis	Yes	Yes	No data
31	SPACA3	Testis	Yes	No	
32	SPATA22	Testis	No	Yes	Yes
33	SSX1	Testis	Yes	No	
34	SSX2	Testis	Yes	No	
35	STK31	Testis	No	No	
36	TKTL1	Testis	No	Yes	Yes
37	TKTL2	Testis	No	Yes	Yes
38	TPTE	Testis	Yes	Yes	Yes
39	TSPYL6	Testis	No	Yes	Yes
40	TUBA3C	Testis	No	No	
41	XAGE1	Testis	Yes	No	
42	XAGE3	Placenta	Yes	No	

Table S4. List of 160 chromatin regulators targeted in the siRNA screen

ARID1A	HDAC2	MBD6	SMARCA1
ASF1A	HDAC3	MECP2	SMARCA2
ASH1L	HDAC4	MEIS1	SMARCA4
ATRX	HDAC5	MPHOSPH8	SMARCA5
AZI2	HDAC6	PBRM1	SMARCB1
BMI1	HDAC7	PCGF1	SMARCC1
BPTF	HDAC8	PCGF2	SMARCC2
BRD1	HMGB1	PCGF3	SMARCD1
BRD2	HMGB3	PCGF5	SMARCD2
BRD3	HOXA5	PCGF6	SMCHD1
BRD4	INO80	PHC1	SUV39H1
BRD7	KAT2A	PHC2	SUV420H1
BRD8	KAT2B	PHF8	SUV420H2
CBX1	KAT6A	PPM1D	SUZ12
CBX3	KAT6B	PRDM2	TDG
CBX5	KDM1A	PRMT1	TET2
CBX6	KDM1B	PRMT5	TET3
CBX7	KDM2A	RBBP4	TRIM24
CDYL	KDM2B	RBBP7	TRIM28
CDYL2	KDM3A	RCOR1	TRIM33
CHD1	KDM3B	REST	UHRF1
CHD3	KDM4A	RING1	UHRF2
CHD4	KDM4C	RNF2	WTAP
CHD6	KDM4D	RXRA	YTHDC1
CHD8	KDM4E	SCAPER	YTHDF1
CTBP1	KDM5A	SCMH1	YTHDF2
CTBP2	KDM5B	SETD1A	YTHDF3
DICER1	KDM5C	SETD1B	ZBTB17
DLX5	KDM6B	SETD2	ZBTB33
DNMT1	KLF5	SETD7	ZBTB38
DNMT3A	KMT2A	SETD8	ZBTB4
DNMT3B	KMT2C	SETDB1	ZBTB40
EED	KMT2E	SETDB2	ZBTB44
EHMT2	LMNA	SIN3A	ZCCHC7
ELP3	LMNB2	SIRT1	ZHX1
EZH1	MBD1	SIRT2	ZHX2
EZH2	MBD2	SIRT3	ZNF114
HAT1	MBD3	SIRT6	ZNF217
HDAC1	MBD4	SIRT7	ZNF416
HDAC11	MBD5	SMAD4	ZNF695

Table S5. List of qRT-PCR primers.

Genes	Forward primers	Reverse primers
ADAM12	AGCTTATGGAACCAAGGAAGAG	CAGTTCTTTGCTTTCCCGTTG
TAK1	CGTCGGAAACCCTTTGATG	CGCTGGGAAGGATCTTTAGAC
KAT2A	CTGTGCTGTCACCTCGAATG	TCGGCGTAGGTTGAGGAAGTA
SIRT6	CAGAGCTCCACGGGAACAT	ACGACTGTGTCTCGGACGTACT
IL6	GCCAGAGCTGTGCAGATGA	ATGTCCTGCAGCCACTGGT
SMAD7	CAGATTCCCAACTTCTTCTGGAG	GTAGAGCCTCCCCACTCTCG
CDKN2B	CGTTAAGTTTACGGCCAACG	GCATGCCCTTGTTCTCCTC
Housekeeping genes		
TBP	TGGCCCATAGTGTCTTTGC	TCCTAGAGCATCTCCAGCACA
PGK1	AGGATAAAGTCAGCCATGTGAG	CACAGGAACTAAAAGGCAGGA

Table S6: Sequence of siRNAs

Genes	Target sequences
KAT2A pool	AGGACAAAUUGGUGCCCGA GUUCCUGGCAUUCGAGAGA GCUACUACGUGACCCGGAA ACUCAUGUCUUUGGGCGAA
KAT2A individual	CCAAGCAGGUCUAUUUCUACC
SIRT6 pool	CCAAGUGUAAGACGCAGUA GUACAUCGCUGCAGAUCCG CCAAAAGGUGAAGGCCAA GAACUGGCGAGGCUGGUCU
SIRT6 individual	CCGGCUCUGCACCGUGGCUAAGG
TAK1 pool	GGACAUUGCUUCUACAAU GAGUGAAUCUGGACGUUUA GGAAAGCGUUUAUUGUAGA GCAAUGAGUUGGUGUUUAC
TAK1 individual	UGGCUUAUCUACACUGGA

Table S7: List of genes in the TAK1 signature

1	ABCA1	CTGF	HBEGF	LRP4	P3H2	SLC35F2	UBASH3B
2	ADAM12	CTPS1	HECTD2	LRRC32	P4HA3	SLC38A5	UBL3
3	ADAM19	DACT1	HIVEP2	MAP3K4	PAG1	SLC46A3	UCK2
4	ADAMTS4	DCBLD1	HMCN1	MAP3K7CL	PDGFA	SMAD7	VAV3
5	AMIGO2	DGKI	HS3ST3B1	MEX3B	PGM2L1	SMURF1	VDR
6	ANGPTL4	DKK1	HTR1F	MFAP3L	PHLDA1	SNAI1	XYLT1
7	ANKRD44	DNAJB5	IER3	MIR181A2	PHLDB1	SNORD56B	ZBTB21
8	APCDD1L	DRP2	IER3	MIR181B1	PKIA	SOCS6	ZNF175
9	ARHGEF40	DUSP6	IER3	MIR181B2	PLEK2	SORBS2	ZNF281
10	ATP10A	E2F7	IFNE	MIR218-1	PLPP4	SPDL1	ZNF365
11	B4GALT1	EDN1	IGF2BP3	MIR221	PLXDC2	SPHK1	
12	BHLHE40	EGR2	IL11	MIR222	PMEPA1	SRPX2	
13	BMP2	ELN	IL6	MIR31	PNMA1	ST6GAL2	
14	BMPR1B	EPHB2	INHBA	MIR31HG	PNP	STARD13	
15	BMPR2	ERBB4	ITGA2	MIR503	PODXL	STK38L	
16	BPGM	ESM1	ITGB6	MIR503HG	PRDM1	SYT14	
17	BTBD11	ETV6	IVNS1ABP	MLXIP	PRICKLE2	TCF4	
18	C4orf26	FAM57A	JADE3	MSC	PTGS2	TGFBI	
19	C5orf46	FAP	JUNB	MURC	RASD2	TMC7	
20	CCIN	FMNL3	KCNJ15	MYOZ1	RASL11B	TMEM2	
21	CDC42SE1	FNDC1	KDR	NCF2	RELT	TMEM51	
22	CDH2	FOXP1	KLF10	NEDD9	RHOB	TNFAIP6	
23	CDK17	FRMD6	LAMC2	NFATC2	SCX	TNFAIP8L3	
24	CHRNA9	FSTL3	LEF1	NNMT	SCX	TNS1	
25	CHST11	FZD8	LHFPL2	NOX4	SEMA7A	TPM1	
26	CNIH3	GALNT10	LIMS1	NREP	SERPINE1	TRIB1	
27	COL4A1	GFPT2	LINC00312	NRP2	SH3PXD2A	TSHZ3	
28	COL4A2	GLIS3	LMCD1	NRP2	SKIL	TSPAN13	
29	COMP	GPAM	LOC79160	OLFM2	SLC19A2	TSPAN2	
30	CPN2	GPR183	LRIG1	OLFML2B	SLC25A32	UACA	

3. Résultats supplémentaires: vers l'étude des cancers

En travaillant sur ce projet, j'ai effectué différentes analyses dont les résultats n'ont pas été publiés, mais qui nous ont aidé à interpréter les données de l'article et dont les retombées m'ont aidées à définir mon second axe de recherche.

Pour valider l'activabilité des gènes C/T dans les IMR90, j'ai analysé des données de RNAseq d'IMR90 traitées par des drogues épigénétiques : la 5-aza-deoxycytidine, un inhibiteur de la méthylation de l'ADN, et la TSA, un inhibiteur de lysine désacétylase (Maunakea 2013). Cette analyse montre que l'inhibition de la méthylation de l'ADN seule ne permet d'activer seulement 6 de nos 42 gènes C/T (*FMR1NB*, *DPEP3*, *TKLT1*, *MAGEA4*, *MAGEA3* et *MAGEA12*), et l'inhibition de la désacétylation n'en active que 2 (*MAGEA12* et *STK31*) des 42 gènes C/T sélectionnés (FIGURE 1)

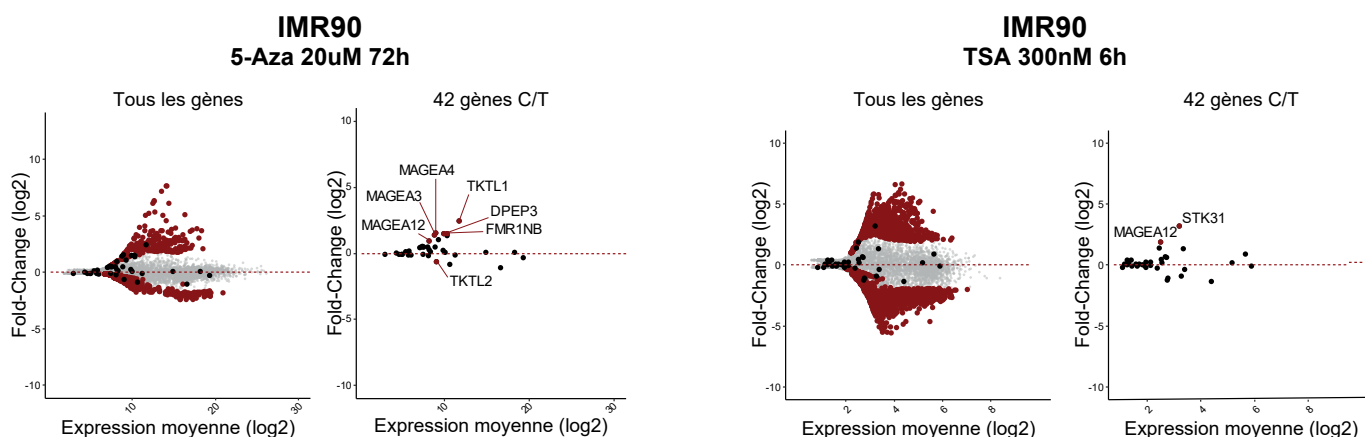


Figure 46 : Réponses transcriptomiques des IMR90 à deux drogues épigénétiques

MAplots représentant les gènes différentiellement exprimés (points rouges), dont certains gènes C/T (points noirs), en réponse à un agent déméthylant (gauche) ou à un inhibiteur de l'acétylation des histones.

Nous voulions nous assurer qu'il existe des conditions pathologiques dans lesquels ces gènes sont activés : pour ce faire, j'ai analysé les données de RNA-seq du TCGA sur les types tumoraux les plus fréquents (poumon, sein et colon : 2576 tumeurs et 263 tissus sains juxta-tumoraux) et identifié les échantillons tumoraux présentant une expression atypique des 42 gènes C/T sélectionnés. Pour faciliter l'analyse, j'ai binarisé les données d'expression en définissant une valeur seuil pour chaque gène C/T, permettant de caractériser son expression comme « atypique » pour toutes les valeurs d'expression au-dessus de ce seuil, et normal sinon (FIGURE 2).

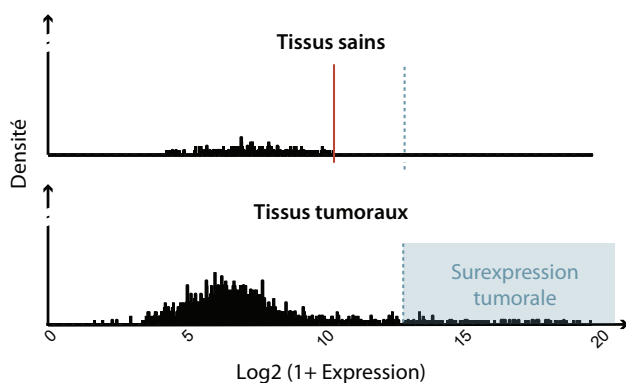
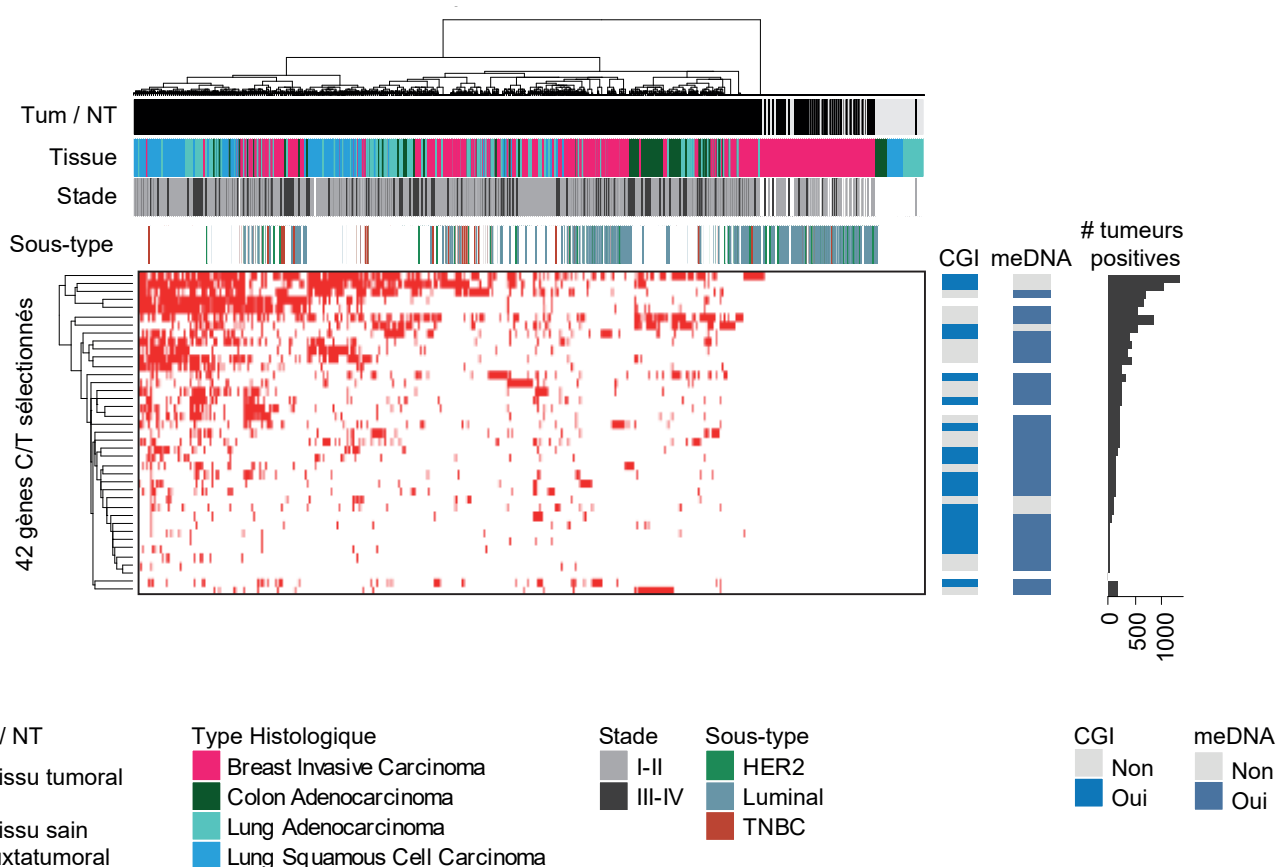


Figure 47 : Représentation schématique des valeurs seuils d'activation des gènes C/T

La valeur seuil est définie pour chaque gène C/T et dans chaque tissu : elle équivaut à la valeur maximale d'expression du gène C/T considéré observée l'organe de référence, auquel on ajoute une fois l'écart-type

J'ai annoté les 42 gènes C/T en identifiant les gènes C/T présentant un îlot CpG dans la région -1500bp / +200bp autour de leur site d'initiation de la transcription ; et en calculant le niveau de méthylation moyen de cette région dans les tissus normaux (colon, poumon et sein confondus).

Les résultats de cette analyse sont résumés dans la **FIGURE 3**. Par construction tous les échantillons de tissus sains juxta-tumoraux sont négatifs pour les 42 gènes C/T testés. Les 42 gènes C/T sont activés dans au moins un échantillon tumoral, bien que les fréquences d'activation soient très variables : le gène le moins fréquemment activé est *ALAS1*, qui n'est exprimé que par un échantillon de tumeur ; le gène le plus fréquemment activé est *ATAD2* (1309 tumeurs). *ADAM12* est exprimé par 560 échantillons, et est le 6^e gène C/T le plus fréquemment activé dans cette cohorte. On notera que les gènes C/T associés à un îlot CpG ont tendance à être moins fréquemment activés que ceux qui n'en possèdent pas. Cependant, on ne parvient pas à définir de cluster clair sur ces 42 gènes C/T.



Néanmoins l'activation des gènes C/T ne semble pas totalement sporadique, car elle permet de définir des clusters parmi les tumeurs. Ces clusters sont organisés en deux groupes : un groupe de tumeurs exprimant un grand nombre de C/T, et comprenant la plus grande partie des tumeurs LUSC et un sous-ensemble des tumeurs LUAD et BRCA ; et un groupe de tumeurs n'exprimant que peu de gènes C/T, et comprenant notamment l'ensemble des tumeurs du colon. Ce résultat est en adéquation avec les caractéristiques publiées de ces types tumoraux vis-à-vis de l'expression des gènes C/T (Hofmann et al. 2008). Plus en détail, on peut identifier des clusters s'organisant par type tumoral : les tumeurs du sein, du colon et du poumon ségrègent plutôt séparément, et on distingue des gènes C/T activés préférentiellement dans un type tumoral plutôt qu'un autre. Au sein des tumeurs du sein, on peut distinguer une tendance à la ségrégation par sous-type tumoral, bien qu'imparfaite à partir de l'expression de ces 42 gènes C/T. Nous aurons l'occasion de revenir sur ce résultat dans la deuxième partie de mon travail.

Afin d'explorer quelques pistes permettant de comprendre pourquoi nos cribles n'ont pu activer qu'ADAM12, et de réfléchir aux caractéristiques génomiques et épigénétiques qui distingueraient éventuellement ce gène des 41 autres gènes C/T, j'ai analysé les données du projet ENCODE : on y trouve des données de RNA-seq et de ChIP-seq sur les IMR90. Ces données révèlent une expression résiduelle d'ADAM12 dans les IMR90, corroborée par la détection de l'ARNpol II sur ce locus (FIGURE 4A). L'expression d'ADAM12 est toutefois très faible en condition basale, si on compare par exemple avec le niveau d'expression d'un marqueur classique des fibroblastes, le gène COL1A1 (contrôle positif : FIGURE 4B). ADAM12 n'est donc pas strictement réprimé dans les IMR90, et se situe probablement dans une chromatine plus permissive. Ceci est confirmé par les données de ChIP-seq, qui montrent un pic significatif en amont du TSS des marques d'histones activatrices H3K4me3, H3K9ac et H3K27ac (bien que très modeste en comparaison avec un gène fortement transcrit comme COL1A1, FIGURE 4B), et l'absence des marques répressives H3K9me3 et H3K27me3. (FIGURE 4A). Concernant la méthylation de l'ADN, ADAM12 présente un îlot CpG dans son premier intron. Les données de MethylArray et de RRBS nous montrent que les CpG analysés dans cette région sont principalement déméthylés, ce qui va dans le sens d'une chromatine plus accessible au niveau du gène ADAM12.

Pour avoir un point de comparaison, on peut regarder les profils d'expression et de modifications épigénétiques deux gènes C/T que nos cribles n'ont pas suffi à activer, l'un présentant un îlot CpG dans son promoteur et l'autre non : MAGEA4 et DPEP3 (contrôles négatifs : FIGURE 4C-D). Tous deux sont activables par un traitement déméthylant global à la 5-Aza-dC (FIGURE 1). Ces deux gènes présentent une absence de signal en RNA-seq et en ChIP-seq dirigé contre l'ARNpol II, indiquant la répression stricte de leur transcription. En cohérence avec leur répression transcriptionnelle, leurs promoteurs montrent un enrichissement visible en H3K9me3 pour MAGEA4 ou en H3K27me3 pour DPEP3. Enfin, les CpG présents dans les domaines promoteurs sont majoritairement méthylés, en particulier pour le gène présentant un îlot CpG dans la région promotrice, DPEP3.

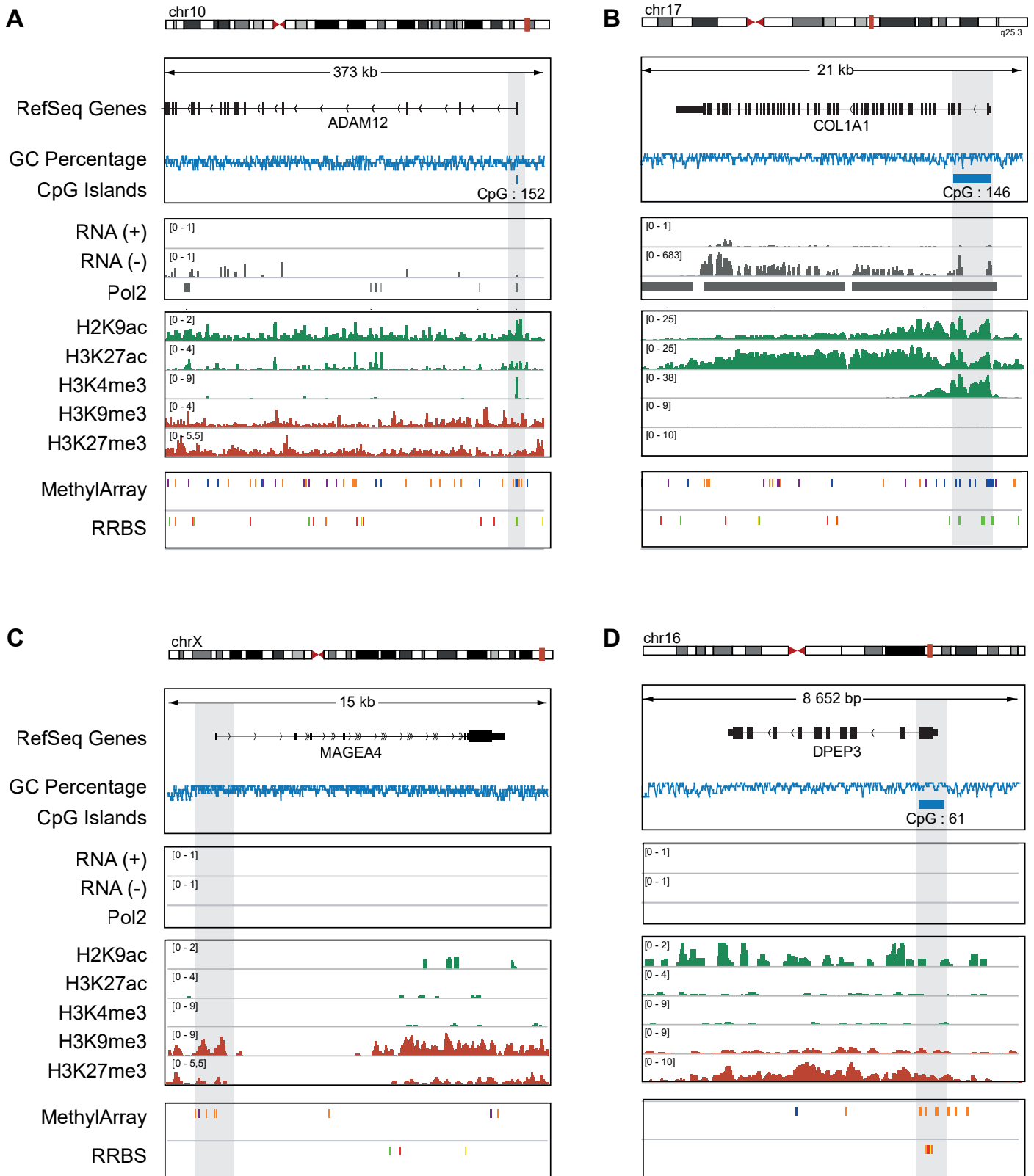


Figure 49 : Paysage épigénétique de 4 gènes dans les IMR90

Visualisation des locus de trois gènes C/T (1, C, D) et d'un gène de lignage fibroblastique (B) dans les IMR90. Ont été inclus dans cette analyse : la définition des îlots CpG, les comptes bruts de RNA-Seq, le résultat d'un ChIP-Seq dirigé contre l'ARNpol II et contre différentes modifications d'histones (représentées en Fold-Change over Control), et des analyses de méthylation par MethylArray (bleu: non méthylé ; orange: méthylé) et RRBS (vert : non méthylé ; rouge : méthylé).

Partie II.

Régulation de l'expression des gènes C/T dans les cancers du sein

1. Définition du cadre expérimental

Dans ce second projet, j'ai cherché à réaliser le chemin conceptuel réciproque : partir de l'analyse des données bioinformatiques sur les cancers, pour formuler des hypothèses qui seront ensuite testées expérimentalement.

Lors du premier projet impliquant ADAM12 et la voie du TGFbeta, j'ai pu généraliser un mécanisme démontré expérimentalement à plusieurs types tumoraux (poumon, colon et sein). Cependant, dans ce second projet, j'ai choisi de me concentrer sur un type de cancers spécifique afin d'alléger la tâche analytique pour qu'elle puisse être réalisée dans le temps imparti. J'ai choisi de m'intéresser aux cancers du sein pour plusieurs raisons :

- D'un point de vue pratique, les cancers du sein sont un des premiers cancers en terme d'occurrence chez les femmes : améliorer la compréhension de cette maladie est donc un enjeu majeur.
- Une conséquence de cette grande fréquence d'apparition des cancers du sein est le volume important de données disponibles : la cohorte de cancers du sein est la plus importante au sein du projet TCGA, et il existe de nombreux jeux de données publics sur cette maladie, permettant d'explorer différents aspects moléculaires et cliniques.
- L'hétérogénéité des cancers du sein est bien caractérisée, et les modèles de classification de ces maladies sont maintenant solidement établis.
- Les cancers du sein sont un cas intéressant du point de vue de la différenciation cellulaire et de l'épigénétique : en effet, comme nous l'avons vu lors de l'introduction, les différents sous-types de cancers du sein proviendraient de la transformation de différentes cellules d'origine. En conséquence, chaque sous-type tumoral émergerait de la combinaison singulière d'un environnement épigénétique et transcriptionnel particulier à sa cellule d'origine, avec les perturbations spécifiques des mutations oncogéniques propre à ce sous-type tumoral. Nous avons donc pensé qu'il s'agissait d'un modèle de choix pour explorer les interactions entre épigénétique et signalisation, comme dans le premier projet.

Lors du premier projet, nous avons appris que le résultat d'un crible dépendait largement du soin apporté au choix du modèle expérimental, c'est pourquoi nous avons passé un certain temps à sélectionner le modèle cellulaire approprié pour tester nos hypothèses. Nous voulions que celui-ci réponde à différents critères :

Nous avons longuement hésité à choisir un (ou plusieurs) modèle(s) de lignées de cancer du sein. Cependant, plusieurs éléments nous ont posé problème. Tout d'abord, les lignées de cancers du sein représentent imparfaitement la diversité des cancers du sein : les tumeurs les plus agressives y sont sur-représentées. Au sein du projet CCLE par exemple, on trouve seulement 20% de lignées de cancer du sein luminaire A, contre 44% de lignées triple-négatives (12% de lumineuses B, 24% de Her2). L'adaptation aux conditions de culture in vitro favorise également la sélection des phénotypes les plus

transformées, et la durée de culture amplifie encore cette dérive génétique et épigénétique : le taux de mutations somatiques par Mb est de 13 dans les tumeurs du TCGA, contre 25 soit le double dans les lignées du CCLE (Jiang et al. 2015). Pour toutes ces raisons, il nous a semblé que des expériences de gain-de-fonction pour l'expression de gènes C/T dans un modèle de cellules de sein peu ou pas transformées nous informerait de façon plus lisible sur les conséquences de ces activations ectopiques pour le développement tumoral, que l'expérience réciproque de perte-de-fonction dans une lignée de cancer du sein, qui présenterait certainement des altérations multiples voire redondantes des voies de signalisation.

Nous avons donc choisi de travailler sur deux modèles cellulaires classiques dans l'étude de la tumorigénèse des carcinomes du sein : les lignées HME et HMLE (FIGURE 1). Ces deux lignées sont dérivées de cellules épithéliales mammaires adultes saines, les HMECs. Elles ont l'avantage de représenter deux étapes intermédiaires de la transformation : les HME ont été immortalisées par l'introduction de la télomérase hTERT ; et les HMLE ont, en plus de l'immortalisation, inactivé les gènes suppresseurs de tumeurs P53 et RB1 grâce à l'introduction du virus oncogénique SV40. L'immortalisation permet de cultiver les HME et les HMLE plus facilement que les HMECs en culture; toutefois aucune de ces deux lignées n'est tumorigénique. En effet, leur injection dans une souris immunodéficiente ne donnera qu'exceptionnellement naissance à une tumeur : pour cela, il manque encore l'activation d'une voie pro-oncogénique, que l'on peut obtenir en infectant les HMLE avec un vecteur codant pour RasV12, ou permettant la surexpression du pro-oncogène MYC par exemple. Les HME et HMLE modélisent donc deux étapes précoces de l'oncogénèse des cancers du sein, et nous ont semblé adapté pour explorer le rôle de l'activation des gènes C/T dans l'acquisition des caractéristiques tumorales.

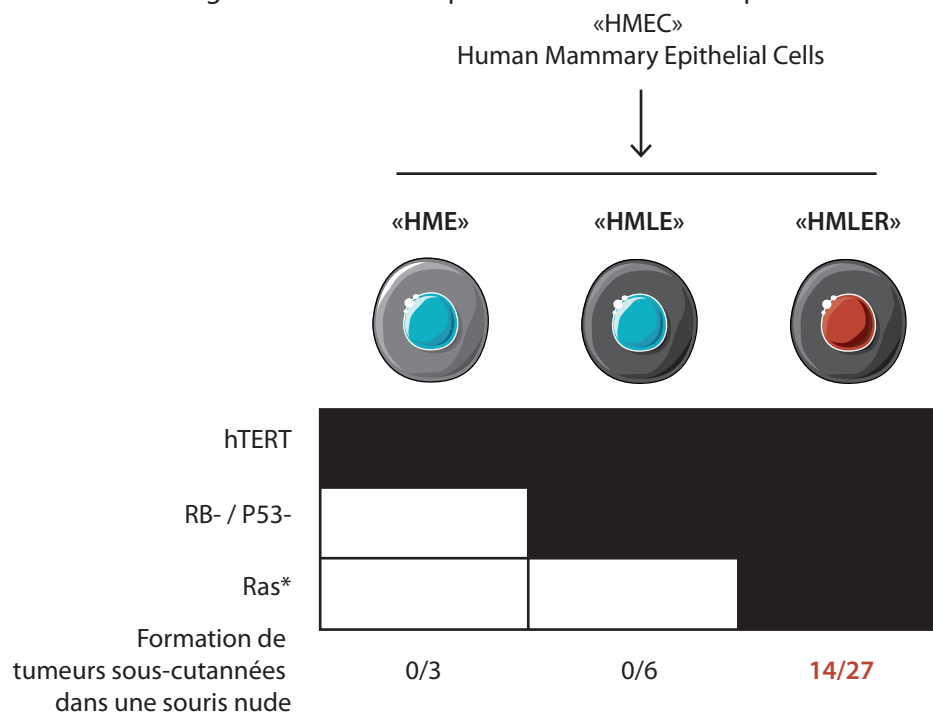


Figure 50 : Modèle cellulaire de tumorigénèse de cellules épithéliales mammaires

Enfin, les HME comme les HMLE sont deux lignées cellulaires hétérogènes, nous permettant de capturer dans une certaine mesure la diversité cellulaire des cellules du sein et d'explorer le rôle de la cellule d'origine dans l'activabilité des gènes C/T. En effet, les HME comme les HMLE sont composées de plusieurs sous-populations cellulaires : la majorité des cellules ont une morphologie plutôt épithéliale, et expriment les marqueurs typiques des progéniteurs luminaux (EpCAM+ CD49f+), et une minorité de cellules sont d'aspect plus mésenchymateux et expriment des marqueurs de cellules souches mammaires (EpCAM- CD49f low/int. et CD44+/CD24 low/neg). Certaines cellules mésenchymateuses ont des propriétés souches plus indifférenciées, et sont capables de générer des descendantes qui se différencieront en cellules épithéliales (**FIGURE 2-3**).

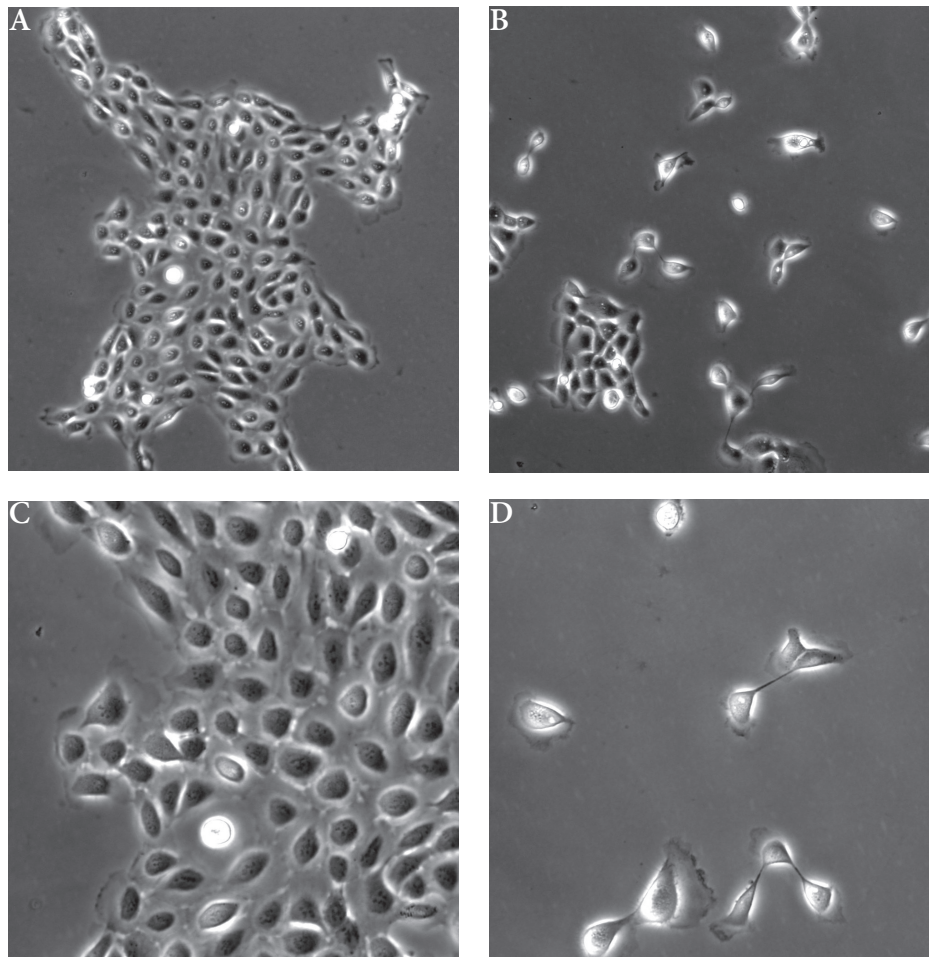


Figure 51 : Phénotype des sous-populations de HMLE par observation microscopique

Ces images ont été obtenues par microscopie à contraste de phase après avoir ensemencé des HMLE à très faible densité et les avoir cultivé pendant une semaine, de telle sorte à obtenir des cellules individuelles se multipliant localement. Panneau B, on observe un îlot de cellules à morphologie épithéliale au sein d'un ensemble de cellules de phénotype plutôt mésenchymateux.

(A, C) : Morphologie épithéliale ; (B, D) : Morphologie mésenchymateuse

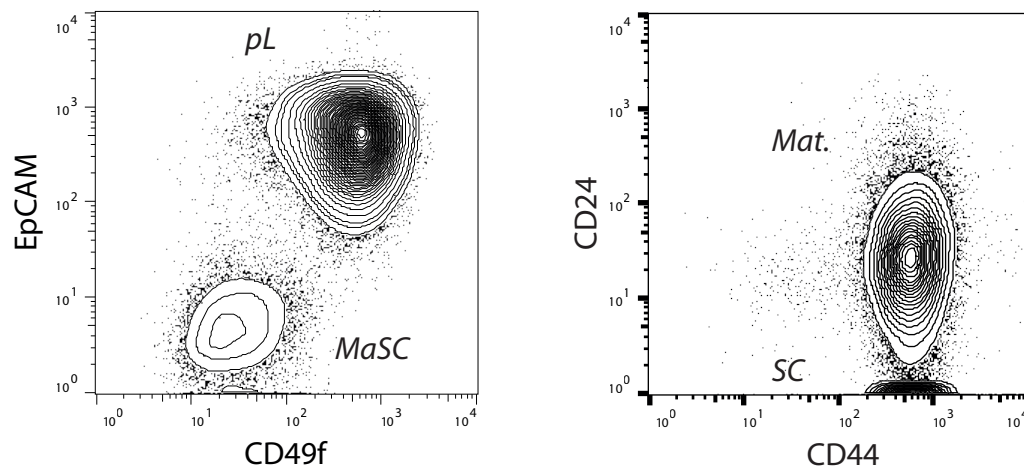


Figure 52 : Phénotype des sous-populations de HMLE par cytométrie

Ces images ont été obtenues par analyse en cytométrie en flux de cellules HMLE, après marquage immunohistochimique. pL: Progéniteurs Luminaux ; MaSC: Mammary Stem Cells ; SC: Stem Cells ; Mat. Mature cells.

Cooperation between Cancer/Testis genes in basal breast tumors

Marthe Laisné¹, Sarah Benlamara¹, André Nicolas², Lounes Djerroudi³, Nikhil Gupta¹, Diana Daher¹, Laure Ferry¹, Olivier Kirsh¹, Claude Philippe⁴, Yuki Okada⁵, Gael Cristofari⁴, Didier Meseure², Anne Vincent-Salomon³, Christophe Ginestier⁶, Pierre-Antoine Defossez^{1*}

Addresses:

¹EDC

²Institut Curie

³Institut Curie

⁴IRCAN, Nice

⁵Tokyo University

⁶CRCM

Abstract

Breast cancer is the most prevalent type of cancer in women worldwide. Within breast tumors, the basal-like subtype has the worst prognosis and no dedicated therapy, therefore new tools to understand, detect, and treat these tumors are needed. Certain germline genes are re-expressed in tumors, and constitute the Cancer/Testis genes; their misexpression has diagnostic and therapeutic applications. Here, we designed a new approach to examine Cancer/Testis gene misexpression in breast tumors. We identify several new markers in Luminal and HER-2 positive tumors, some of which predict response to chemotherapy. We then use machine learning to identify the 2 Cancer/Testis genes most associated with basal-like breast tumors: *HORMAD1* and *CT83*. We show that these genes are expressed by tumor cells but not the microenvironment, and that they are not expressed by normal breast progenitors, in other words their activation occurs de novo. We find these genes are epigenetically repressed by DNA methylation, and that their activation upon DNA demethylation is irreversible, providing a memory of past epigenetic disturbances. Basal-like tumors expressing both genes have a poorer outcome than tumors expressing either gene alone or neither gene. Expression of both genes together in breast cells in vitro has a synergistic effect, increases stemness, and activates a transcriptional profile also observed in double-positive tumors. Therefore, we reveal a functional cooperation between Cancer/Testis genes in basal breast tumors; these findings have consequences for the understanding, diagnosis, and therapy of the breast tumors with the worse outcomes.

INTRODUCTION

Cancer cells undergo massive genetic and epigenetic changes relative to their normal progenitors. The advances of genomics and epigenomics have yielded an ever more complete picture of these abnormalities, and drawn accurate molecular portraits of different tumor types. The large number of samples examined in public cohorts increase statistical power, yet parsing out driver from passenger events remains far from trivial (Muñíos F. et al., 2021).

Altered gene expression is one of the functional consequence of genetic and epigenetic modifications in tumors. Genes can be turned off by deletions, alterations in their control elements such as enhancers, or changes in the transcriptional machinery. Conversely, they can become overexpressed by amplification, gain of enhancers, or expression of transcriptional activators, among other possibilities. Genes that are frequently turned on in a tumor type are useful as biomarkers. In some instances, their expression can inform prognosis and choice of treatment. Finally, these overexpressed genes can play a physiological role in the tumor cells, and therefore represent therapeutic targets. HER2 is such an example: the gene can be amplified, its overexpression marks a specific subtype of breast tumors, and highly efficient therapeutic antibodies have been generated against this target.

HER2 is expressed by normal breast cells, so its overexpression in breast tumors is just the amplification of a pre-existing expression pattern. However, tumor cells can also deviate radically from their ancestral gene expression pattern and turn on genes that are normally activated in other tissue types or other developmental stages (Wang J. et al. 2014). For instance, various tumor types, in men and women, express genes that are typical of the placenta (Rousseaux S. et al. 2014; Naciri et al. 2019). Within this broad framework of ectopic gene reactivation in tumors, one class of genes bears special conceptual interest and therapeutic promise: the cancer/testis genes.

As their name implies, the cancer/testis genes are normally expressed only in the male germline, but become reactivated in tumors, both in female and male patients (Whitehurst AW 2014). As they are not expressed in any normal somatic cells, they are remarkable biomarkers for tumors. In addition, as the testis is an immune sanctuary in men, and as the testicular genes are not normally expressed in women, their expression in tumors opens an excellent possibility for immunotherapy. Finally, cancer/testis genes may be oncogenes in their own right, and are potential drug targets for therapy (Gibbs ZA & Whitehurst AW 2018).

Breast cancer is the most common cancer in women, both in developed and developing countries, and breast malignancies killed almost 700,000 women worldwide in 2020 (www.who.int). It has long been appreciated that breast tumors form an heterogeneous ensemble, with at least 5 distinguishable subtypes: normal-like, Luminal A, Luminal B, HER2-positive, and basal-like. Within those groups, basal-like tumors could themselves contain distinct subtypes, and they have the worst prognosis and no dedicated therapy.

Cancer/testis genes have been investigated as potential biomarkers, oncogenes, and targets in breast cancer, with promising results (Kaufmann J. et al. 2019; Paret C. et al. 2018; Adams S. et al. 2011; Mischo A. et al. 2006). To build on these investigations, we undertook an unbiased analysis of publically available expression data with a new bioinformatic approach. This led us to discover several new markers associated with different breast tumor subtypes. Our cohort of in situ tumors establishes that cancer/testis gene activation is an early event in tumorigenesis, and that there is no switch of their expression pattern between early and more established tumors. We then focused on the two genes whose expression is most highly associated with basal breast tumors: *HORMAD1* and *CT83*. We show that these genes are not expressed by healthy progenitors, but expressed *de novo* in the tumor cells. We demonstrate that loss of methylation is sufficient to reactivate both genes, and that an initial activation event is sufficient to trigger persistent expression. Most basal tumors express at least one of the two genes, but those that express both have significantly worse outcome, hinting at a cooperative effect. Using breast cells in culture, we prove that the two genes synergize to modulate stemness and initiate a transcriptional signature that is also found in basal tumors expressing *HORMAD1* and *CT83* simultaneously. These findings advance our conceptual understanding of cancer/testis genes in breast cancer, and they have practical implications for diagnosis and treatment.

RESULTATS

A custom bioinformatic approach identifies the Cancer/Testis genes most associated with breast tumors

The first step of our study was to establish an exhaustive list of C/T genes; it includes all of the C/T genes described in three independent publications, for a total of 1350 genes (Almeida et al. 2009; Rousseaux et al. 2013; Wang et al. 2016). Our second resource was genomics data, including RNA-seq, from The Cancer Genome Atlas (TCGA), covering 1090 tumors samples and 113 healthy juxtatumoral mammary samples (Figure 1A).

To identify C/T genes reactivated in breast tumors, we established a custom bioinformatic approach. An ideal biomarker should have little or no expression in healthy samples, but high expression in at least some of the tumors. Mathematically, these properties are reflected in a zero-centred, single-mode density function in healthy breast samples, and a multi-mode density function with one or more non-zero maxima in tumor samples, reflecting one or more groups of tumors that have activated this gene. Such profiles can be detected automatically by examining changes in the derivative of the density function (Figure 1A).

To implement this idea, we created a two-step pipeline. First, we determined the distribution of expression of each C/T gene in healthy mammary samples and in breast tumors, and smoothed these distributions using kernel density estimation. As it is crucial to not overfit or oversmooth expression values, we systematically tested multiple values for the bandwidth parameter using positive and negative controls (Figure S1A) and we selected a balanced value (bandwidth = 0.7). Second, we analyzed the derivative of the distribution function to obtain the number of distinct peaks. This allowed us to focus on C/T genes that are not expressed in healthy mammary samples (unimodal expression profile centered on 0 according to kernel density estimation), but activated in some breast tumor samples (multimodal expression profile).

Our method complements previously used approaches in that it is orthogonal, less calculation-intensive, flexible, and sensitive. Of note, this unbiased scheme is not restricted to C/T genes and it could be broadly used to identify any other genes that are abnormally expressed in tumor samples compared to matched normal juxta-tumor tissues, such as potential tumor suppressor genes or oncogenes (Figure S1B-D). Our approach allowed us to define a highly selective list of 139 C/T genes with abnormal expression profile in breast tumors compared to normal breast (Figure 1B, Supplementary Table 1). The examination of GTEx RNA-seq data confirmed that these 139 genes are expressed in the human germline, but not in the breast (or other healthy tissues, Figure S1G). The reactivation seen in tumors is therefore a pathological event.

The activation of selected C/T genes marks different subtypes of tumors and cell lines

We then tested whether the expression of certain members of our 139-gene list was specifically associated with certain subtypes of breast tumors. For this, we used Principal Component Analysis (PCA) on TCGA data, using the subtype annotations provided for each tumor (Figure 2A). A visual inspection suggested that tumor types could be separated on the basis of C/T gene expression (Figure 2A), with a clearly distinct group of basal tumors, for instance. These clusters were also found when the tumors were classified on the basis of their anatomohistological subtype, rather than their transcriptome-defined subtype (Figure S2A), and they were also visible when UMAP was used instead of PCA (Figure 2A, S2A). We therefore conclude that expression of some genes in our list can stratify breast tumors by subtypes.

To identify these genes systematically we used a machine learning approach. We established a random forest model on a training set of TCGA breast tumors (75% of all samples, n=817), and tested the best model on the remaining tumors (n=273). This model could very effectively identify basal tumors, with high sensitivity (0.9) and high specificity (1.0), leading to a balanced accuracy nearing 100% (2B). Again, similar results were found when the tumors were classified anatomopathologically, rather than transcriptionally (Figure S2B). For Luminal B and Her2 subtypes the specificity

scores were high (1.0 and 0.9 respectively), but the sensitivity lower (0.4 and 0.2) (Figure 2B). This could be due to the fact that some tumors of these groups do not express any C/T genes, leading to a lack of available information for the prediction.

Using the best random forest model, we ranked the 139 C/T genes according to their predictive value; the top 15 C/T genes are depicted in Figure 2C (and in Figure S2C for the analysis carried out with anatomopathological stratification). The two best predictors, *HORMAD1* and *CT83*, are strongly associated with basal breast tumors: of the 190 basal-like breast tumors, 89% expressed either *HORMAD1* or *CT83*, compared to only 13% of HER2-amplified, 6% of Luminal B, and 2% of Luminal A tumors (Figures 2D and S2D). These results are consistent with several previous reports that have associated *HORMAD1* or *CT83* expression with basal tumors (Watkins et al. 2015; José Adélaïde et al. 2007; Chen et al. 2019; Paret et al. 2015; Kondo et al. 2018; Chen et al. 2021), and they validate our approach. *HORMAD1*, a gene on human chromosome 1q21.3, is physiologically expressed by the pre-leptotene spermatocytes (Shin et al., 2010) and it regulates meiotic progression. *CT83*, on the other hand, is located on human chromosome region Xq23, it is expressed in mature sperm (Jung et al., 2019) but its reproductive function is unknown.

The expression of two other markers, *DMRTC2* and *TDRD1*, is associated with HER2-positive tumors (Figure 2D), but the association is looser than that of *HORMAD1/CT83* with basal tumors. During spermatogenesis, *DMRTC2* has essential functions during pachytene (Date et al. 2012), whereas *TDRD1* interacts with piRNAs and Piwi proteins to promote silencing (Mathioudakis et al. 2012). To the best of our knowledge, neither *DMRTC2* nor *TDRD1* have been previously linked to breast cancers in general, and to the HER-2 positive subtype in particular.

Lastly, we found two markers, *LRGUK* and *TEX14*, for which expression tends to mark Luminal tumors (Figure 2D). *LRGUK* is involved in diverse aspects of sperm assembly, including the microtubule-based shaping of spermatozooids (Liu et al. 2015); it was more frequently over-expressed in luminal A breast tumors (Figure 2D). As for *TEX14*, a factor necessary for intracellular bridges in germ cells (Greenbaum et al. 2006), it marked luminal B breast cancers, as well as luminal A tumors to a smaller extent (Figure 2D). While *TEX14* has previously been linked to basal breast tumors (Karlin et al. 2015), we believe we present the first report that is actually much more prevalently expressed in Luminal tumors, especially of the more aggressive B subtype, and we are not aware of any publications linking *LRGUK* to breast tumors in general, nor to Luminal tumors in particular.

We next tested whether the associations we had detected using tumor expression data also held true with cancer cell lines. For this, we determined the expression level of the 6 markers described above in all the breast cell lines found in the Cancer Cell Line Encyclopedia (Figure S2E). We observed a good general agreement between tumors and cell lines of the same subtype. For instance, *HORMAD1* and/or *CT83* were highly expressed in the basal cell lines such as MDA-MB-436, MDA-MB-468, and HCC1599, but not in Luminal or HER2-positive cells. *DMRTC2* and/or *TDRD1* expression marked HER2-positive lines like AU565 or SKBR3. Finally, a typical Luminal A line, MCF7, expressed *LRGUK* and *TEX14*.

Expression of the markers can be detected in early-stage tumors of the relevant subtype

The 6 markers we report were found in an unbiased analysis of the TCGA breast tumor set, which contains mostly mid- and late-stage malignancies. A question that is important practically and conceptually is whether these markers are already expressed at early stages of tumorigenesis. To answer this question, we used RNA-seq analysis of early tumors (*in situ* and micro-invasive) and invasive breast carcinomas of different subtypes (n=55, our unpublished INVADE cohort, Figure 2E). Twenty-four of the 35 early tumors (68%) expressed at least one of the markers, while 11 out of 20 invasive tumors (55%) did so. The association between marker and tumor type was generally respected: for instance *LRGUK* was expressed in 14 tumors, of which 11 were luminal (p-value= $2 \cdot 10^{-7}$), seven of those being early-stage and the remaining four invasive. *TDRD1* was expressed in 12 samples, of which 7 were HER2-positive (p-value= $2 \cdot 10^{-4}$), and 6 out of those 7 were early-stage. *DMRTC2* was not found in any of the early HER2-positive samples, possibly indicat-

ing that its expression is induced later in tumorigenesis. The expression of *HORMAD1* and *CT83* was rare, which is not surprising as basal tumors are rarely found at early stages. To summarize, we can draw several conclusions from this data set: 1) the activation of C/T genes can be an early event during tumorigenesis, detectable on in situ tumors; 2) the type of C/T genes activated in a tumor is not different between early and later-stage tumors: there is no switch in expression. These conclusions are further substantiated by analyzing the same dataset with every C/T gene expressed in at least one tumor of our cohort (39 genes, Figure S2F).

Marker expression can be associated with response and survival

Finally, we asked whether the expression of these CT genes could distinguish, within a breast cancer subtype, tumors with a different prognosis or therapeutic response. We examined relapse-free survival at more than 10 years, on a large panel of breast tumors of known subtype (Györfy 2021).

Activation of *LRGUK* in Luminal A or Luminal B tumors, was an indicator of good prognosis (Figure 2F). Furthermore, activation of the gene tended to correlate with better response to anthracyclines, although the trend failed to reach significance (Figure 2G).

For Her2-positive tumors, the expression of *TDRD1* was not statistically linked to survival, whereas *DMRTC2* expression correlated with poorer survival (Figure S2G). To detect other potentially useful characteristics of these tumors, we examined their immunological signature with the Immunoscope tool (Bindea et al. 2013) (Figure S2H): those with high *DMRTC2* were more “hot”, i.e. more infiltrated, but also more immunosuppressive (high *FOXP3* activation). Therefore, they might be attractive candidates for treatment with immune checkpoint inhibitors (Galon et al. 2019). As far as we are aware, all of these associations are new and may be helpful for prognosis and treatment choice.

The situation was particularly interesting for *HORMAD1* and *CT83* in basal-like tumors (Figure 2G). Neither gene considered alone was associated with prognosis, however the co-expression of both genes led to a significantly worse outcome, hinting at a possible synergistic effect. In addition, expression of both genes simultaneously correlated with a poorer response to anthracycline chemotherapy (Figure 2I).

***HORMAD1* and *CT83* mark are expressed by most cancer cells in basal-like tumors, but are not expressed by the microenvironment**

As basal-like tumors are especially deadly, we aimed the rest of our investigations on this tumor type. We started by repeating our random forest analysis on RNA-seq data from an independent set of tumors (Varley et al. 2014). In that second cohort also, *HORMAD1* and *CT83* were the most informative genes, and the most associated with basal tumors (Figures 3A and 3B). This independent cohort further supports the relevance of these 2 genes in basal tumors, thus we focused on *HORMAD1* and *CT83* in the rest of our work.

In the TCGA cohort, ~90% of basal-like tumors expressed *HORMAD1* or *CT83* at the RNA level, and ~60% expressed both (Figure 3C). Basal-like tumors are a heterogeneous ensemble, but tumors expressing both *HORMAD1* and *CT83* tended to form a more homogeneous set, with fewer distinct anatomopathological groups and a reduced number of molecular signatures (Figure S3A, Supplementary Table 2). Using the Lehmann classification (Lehmann et al. 2016), we found double-positive tumors in all subgroups except for Luminal Androgen Receptor (Figure S3B). In breast cancer cell lines as well, 70% of basal-like cell lines from CCLE were positive for *HORMAD1* and/or *CT83* (Figure S3C).

We then sought to confirm and complement these transcriptional analyses with immunohistochemistry (IHC). We screened antibodies and experimental conditions until we arrived at combinations under which the IHC pattern observed on human testis sections matched the results of single-cell RNA-seq in the same organ (Sohni et al. 2019). With these conditions, we could observe nuclear staining for *HORMAD1* specifically in preleptotene spermatocytes,

and staining in mature spermatozooids for CT83 (Figure 3D). Using the same conditions on 99 tumor sections of mixed types, we verified that most triple-negative tumors (34 out of 40, 85%) expressed HORMAD1 and/or CT83. Conversely, HORMAD1 and CT83 were significantly more often detected in triple-negative breast tumors (Figure S3D: p -value $< 10^{-4}$). In the positive tumors, staining for HORMAD1 was predominantly nuclear, present in most or all tumor cells, and seemed absent from non-tumor cells of the microenvironment. CT83 staining was cytoplasmic, but similarly marked most tumor cells, and few or no cells of the microenvironment (Figure 3D).

To verify these findings with an orthogonal approach, we re-analyzed previously published single-cell RNA-seq data of 6 triple-negative breast tumors (of which 5 express HORMAD1 and CT83) (GSE75688, Chung et al. 2017). We found very clearly that only tumor cells (and not the microenvironment) express HORMAD1 and/or CT83 (Figure 3E). Within any given tumor, approximately 20-40% of individual cancer cells express either HORMAD1 or CT83, and around 5-20% express both.

Taken together, these results at the RNA and protein level show that HORMAD1 and CT83 are expressed by most tumoral cells in most basal-like tumors, and that they are not expressed by the microenvironment.

Most healthy mammary cells fail to express HORMAD1 or CT83

As HORMAD1 and CT83 are expressed by tumor cells, and as these tumor cells derive from the transformation of healthy breast cells, we asked whether the 2 genes are expressed by progenitors found in healthy breast. For this, we turned to RNA expression data obtained on healthy cells sorted from reduction mammoplasties, where markers were used to FACS-sort stem cells, luminal progenitors, and mature luminal cells (Figure 3F, Morel et al. 2017). Known genes displayed the expected expression pattern: for example MSRB3 was expressed in stem but not more differentiated cells, whereas ESR1 had the opposite pattern (Figure 3G). In contrast, neither HORMAD1 nor CT83 was detectably expressed in any of the sorted cell populations (Figure 3G). In particular, they were not detectably expressed in luminal progenitors, which are the proposed cells of origin for basal tumors (Molyneux et al. 2010). Therefore, expression of CT83/HORMAD1 in basal tumors does not seem to merely reflect pre-existing expression in the cells of origin of the tumors.

We investigated this question further using single-cell RNA-seq data from normal human breast. Using a combination of dimensional reduction, unsupervised clustering approaches, and previously known markers, we were able to separate the luminal from the basal-epithelial compartments (Figure 3H). The expression of MSRB3 and ESR1 marked the expected populations (Figure S3E). We detected some normal cells expressing CT83 and/or HORMAD1 (Figure 3H, red circles), however these cells were very rare: only 15 out of 24 292 total cells expressed HORMAD1 and/or CT83. The positive cells either that could be assigned to a cluster were mostly “Luminal Epithelial” cluster, however more than 99% of Luminal Epithelial cells failed to express HORMAD1 or CT83, which is consistent with the lack of detection in the sorted cell populations of Figure 3G.

Expression of HORMAD1 and CT83 in tumors correlates with promoter demethylation

Basal-like tumors are genetically unstable (Russnes et al., 2017), so we examined whether HORMAD1 and CT83 overexpression could be due to gene amplification. We found two results arguing against this possibility. First, there were no correlations between Copy Number Variation (CNV) and mRNA levels for HORMAD1 or CT83 in basal tumors (Figure 4A). Second, if the genes were overexpressed because their locus is amplified, then we would expect to see a positive correlation between the expression of HORMAD1 and its two adjoining genes (GOLPH3L, 1kb away, and CTSS, 9 kb away), and/or between CT83 and its contiguous gene SLC6A14 (250 base pairs away). We failed to detect any such correlation, whereas the expression of a gene known to undergo amplification and used as a positive control in the

analysis, ERBB2, correlated positively with the expression of the neighboring gene PGAP3 (Figure 4B).

As amplification seemed unlikely to explain the overexpression of *HORMAD1* and/or *CT83*, we next examined epigenetic events. The genes lack CpG islands, but both have promoters with an intermediate CpG density (ICP) (Figure 4C). These promoters overlap ATAC-seq peaks that are present in *HORMAD1/CT83*-expressing breast tumors, but absent in non-expressing tumors (Figures 4C and S4C). We next investigated the DNA methylation status of these promoters, using the Illumina 450K arrays available in TCGA and GEO. As shown in Figure S4B, we found high levels of methylation on the *HORMAD1* and *CT83* promoters in normal breast samples (that do not express the genes) and low levels of methylation in the sperm samples (where the genes are on). The data in tumors show a very strong correlation between expression and promoter demethylation for *CT83* (Figure 4D). The correlation is present but less absolute for *HORMAD1*, as some tumors overexpress *HORMAD1* without displaying demethylation. These specific tumors tend to have a higher *HORMAD1* copy number (Figure 4D), and our hypothesis is that most of the copies are methylated and silent, while a few are demethylated and active.

We then tested functionally whether demethylation suffices to induce *HORMAD1* and *CT83* expression. For this, we used immortalized human mammary epithelial cells (HME and HMLE, Elenbaas et al. 2001) treated *in vitro* with 5-aza-deoxy-cytidine (5-aza-dC). The treatment induced both genes, in a dose-dependent manner (Figure 4E), and led to detectable protein expression (Figure 4F). Importantly, the genes remained expressed even after the drug was removed (Figure 4G), demonstrating a memory effect.

To better characterize the epigenetic landscape of *HORMAD1* and *CT83* in both normal and pathological conditions, we used public ChIP-seq datasets. In the testis, *HORMAD1* showed a significant enrichment in the activating histone marks H3K27ac and H3K4me3, which were absent in breast. Conversely, in the breast, *HORMAD1* and *CT83* were marked by the repressive chromatin mark H3K9me3 (Figure S4A). The activation marks H3K27ac and H3K4me4 were also found for *HORMAD1* and *CT83* in the basal-like breast cancer cell line MDA-MB-436; but surprisingly we did not detect repressive marks in the non-tumorigenic mammary cell line MCF10A nor in the luminal A breast cancer cell line MCF7 (Figure S4B). From these data we conclude that *HORMAD1* and *CT83* are normally silenced by DNA methylation and, likely, H3K9me3 methylation, and that these marks are lost and replaced by active modifications such as H3K4me3 in cell lines and tumors that re-express the genes.

HORMAD1 and CT83 act synergistically to increase stem-like cell proportions *in vitro*

HORMAD1/CT83 expression may just be a bystander consequence of epigenetic instability, or it could have a positively selected function; in other words the genes and their products could be either markers or actors. To investigate this question experimentally, we used the HMLE cells (human mammary epithelial cells expressing hTERT and large T/small T), which constitute a well-accepted model to study the genesis of basal-like tumors. We generated polycistronic lentiviral vectors to express *HORMAD1* and/or *CT83* and select the infected cells with antibiotics (Figure 5A). RNA and protein were expressed as expected by the different vectors (Figures 5B-5C). Importantly, these first experiments showed that expression of one gene was not sufficient to induce the other; this is consistent with our observation that expression of *HORMAD1* or *CT83* alone is not equivalent to expression of both genes. By immunofluorescence with the cognate antibodies, we confirmed the published nuclear localization of *HORMAD1*, and a perinuclear localization for *CT83* that could correspond to the endoplasmic reticulum (Figure 5D), these patterns matched those observed with GFP-tagged proteins (S5A). The expression of one protein (*HORMAD1* or *CT83*) did not measurably affect the distribution of the other (5D), suggesting that they function independently of each other, in different cellular compartments.

Because the simultaneous activation of *HORMAD1* and *CT83* is associated with a worse prognosis for basal-like tumors, we carried out functional experiments to examine the consequences of expressing one gene, the other, or both at the same time. We first simply examined growth rate, and observed no significant difference between control cells, and

cells expressing HORMAD1 and/or CT83 (Figure S5C). Therefore there does not seem to be a positive selection for accelerated proliferation in cells expressing both HORMAD1 and CT83.

We then examined other cellular phenotype relevant to tumorigenesis: cellular identity and stemness. HMLE cells grown in vitro maintain some of the heterogeneity and differentiation hierarchy of the mammary gland. FACS sorting using well-characterized markers (CD49f/EpCAM) showed, as expected, that the control cell population contained ~95% of Luminal Progenitor-like cells (LP), about 3% of mammary stem cell-like cells (MaSC), and about 1% cells resembling Mature Luminal (ML) cells (Figure 5E). We therefore measured the proportion of each population after expression of HORMAD1, CT83, or both. Expression of HORMAD1 alone tended to decrease the percentage of MaSC, but this trend failed to reach statistical significance. Expression of CT83 alone failed to elicit detectable variations. In contrast, the co-expression of HORMAD1 and CT83 induced a highly significant, 4-fold increase of the MaSC compartment (Figure 5E). This finding was confirmed using a different set of markers, CD24/CD44; there again, we saw an increase of the CD44-positive, CD24-low compartment, corresponding to stem-like cells (Figure 5F). Finally, we also observed the synergistic effect of HORMAD1 and CT83 on stem-like cells in a second cellular model, the HME cells (Figure S5 D-F). We then sought to better understand the molecular bases of this synergy.

Simultaneous HORMAD1 and CT83 expression triggers a transcriptional signature found in tumors

We next performed an RNA-seq experiment in HMLE cells expressing HORMAD1 and/or CT83. The expression of HORMAD1 or CT83 alone had no major impact on the transcriptome, but co-expression of the 2 genes did induce a specific gene signature, with 88 differentially expressed genes, 49 down and 39 up (Figure 6A). Among these, we found genes known to be associated with basal breast cancer, such as ETV1, PGHDH, and LMO4 (induced), or ELF3, FOXO3, and PLK2 (repressed). We then searched for biological functions associated with this “HORMAD1+CT83” signature. We performed Gene Ontology (GO) analyses on curated signatures or general pathways. We found significant associations of this signature with several breast cancer-related pathways, and with pathways associated with epithelial to mesenchymal transition and specific signaling pathways of the mammary gland (Figure 6B). We verified by qRT-PCR the RNA-seq results on selected genes (Figure S6A). The decrease of E-Cadherin (CDH1) following joint HORMAD1/CT83 expression was also detected in HME cells (Figure S6B).

We subsequently asked whether the “HORMAD1+CT83” signature, as seen in vitro, is germane to the transcriptional profile of cancer cell lines and of basal tumors expressing both genes. For this, we started by performing differential gene expression analysis on the double-positive vs. double-negative basal cell lines within the CCLE. This yielded a transcriptomic profile including 440 differentially regulated genes (Figure S6C). A GSEA analysis showed that the “HORMAD1+CT83” signature detected in HMLE cells was significantly correlated to the transcriptome profile of double-positive cancer cell lines (Figure 6C).

Next, we carried out a similar analysis on the basal-like tumors in the TCGA. The transcriptomic profile of HORMAD1/CT83 double-positive tumors was different from that of double-negative tumors and included 571 differentially expressed genes (Figure S6D). Again, the in vitro “HORMAD1+CT83” signature significantly correlated with the transcriptome of double-positive tumors (Figure 6C).

Together, these results establish that expression of HORMAD1 and CT83 together in breast cells has effects different from the expression of either gene alone. In addition, the expression of both genes is sufficient to induce a transcriptional program that resembles the

DISCUSSION

A new approach identifies cancer/testis genes expressed in different breast tumor subtypes

Cancer/Testis genes hold promise as markers, actors, and targets in cancer. Here we implemented a new bioinformatic approach to identify the Cancer/Testis genes that are overexpressed in breast cancer. This approach has the advantage of being rigorous and calculation-efficient, immediately usable for any tumor type, but also easily adaptable to seek other types of genes misexpressed in tumors. It complements previous approaches based on expression thresholds (Rousseaux et al. 2014) or vector colinearity (Wang et al. 2016), and yielded results that either approach alone would not have yielded (Figure 1B).

This approach, combined with machine learning on large breast cancer cohorts, has led us to uncover new markers that are specific of different breast cancer subtypes. Most of them were previously unknown, and some of them are associated with prognosis and response to treatment: they may become valuable markers. In addition, future investigations could examine whether they actively participate in the transformation process. Our examination of early-stage tumors reveals that the pattern of cancer/testis genes expression is determined early on, which has interesting practical and conceptual implications.

We identify two genes —CT83 and HORMAD1— that are expressed by most basal tumors, but few other tumor of the other subtypes. By definition, these genes are normally expressed in the testis. HORMAD1 is expressed in pre-leptotene spermatocytes (Shin et al., 2010), it is required for the promotion of non-conservative recombination events in meiosis and the resulting formation of the synaptonemal complex (Kumar et al. 2015). CT83 (also known as CXorf61 or KK-LC-1) encodes a small protein (113 AA) of unknown function, normally expressed in mature sperm (Jung et al. 2019).

Both genes had been previously linked to basal tumors (Holm et al. 2016; Kaufmann et al. 2019; Wang et al. 2018; Watkins et al. 2015; Zhong et al. 2020), but our work goes further and brings a number of novel findings : 1) we rigorously prove that the genes are the 2 strongest predictors of a tumor being basal in independent cohorts, 2) we show that the genes are not expressed in healthy breast progenitors, showing that the induction occurs de novo, 3) we demonstrate that they have a synergistic effect on breast cells.

Three important questions remain open and will be discussed briefly in the following paragraphs: what is the order of events leading to HORMAD1/CT83 induction in basal tumors? What are the mechanistic bases for their induction? And how do the two genes exert their synergistic effect?

Order of events

About 90% of basal tumors in the TCGA cohort express HORMAD1 or CT83, and about 60% express both. There are two non-exclusive interpretations for these high proportions.

First, the induction of the genes could be an early event that occurs in most early lesions and is maintained as the tumor progresses. In principle, this deregulation could even occur earlier than the main transforming event, such as activation of Myc. It could be that HORMAD1/CT83 induction reflects a disturbed epigenetic landscape in rare tumor-initiating cells, which could itself increase the probability of cellular transformation. In that possibility, HORMAD1 and CT83 themselves could just be markers of the early epigenetic instability, or they could actively participate in the ensuing transformation. One piece of data supporting this “induction before transformation” hypothesis is that a few rare cells in the healthy breast already express CT83 and/or HORMAD1. Some of those aberrant cells might eventually be amenable to enter the basal-like transformation path.

Second, it could be that the expression of both HORMAD1 and CT83 occurs after transformation and brings a selective

advantage to basal tumor cells. The genes have only been studied individually so far, but there is convincing evidence that *HORMAD1* overexpression impairs homologous recombination and increases genomic instability in basal breast tumor cells, therefore possibly speeding up tumor evolution (Watkins et al. 2015). *HORMAD1* overexpression is also detected in lung tumors but, paradoxically, it seems to increase the robustness of homologous recombination in these tumors, making them more resistant to DNA-damaging chemotherapy. These divergences may mean that *HORMAD1* has context-dependent functions, for instance in the presence or absence of other actors such as *CT83*.

Mechanism of induction

While basal tumors are genetically unstable, we rule out gene amplification as the main mechanism of *HORMAD1/CT83* induction. Instead, we show that DNA methylation is a barrier to *HORMAD1/CT83* activation, which is consistent with previously published reports (Nichols et al. 2018; Wang et al. 2018; Chen et al. 2019). Importantly, we find that, once the genes have been induced by a 5-aza-deoxycytidine treatment, they remain active even when 5-aza-dC has been removed. In other words, they switch to a stable “On” state. This makes them excellent markers of past epigenetic disturbances.

Further investigations will be required to elucidate the initial event(s) that lead to the derepression of *HORMAD1/CT83* at some point during the history of most basal tumors. It could be a stochastic phenomenon occurring before or after transformation; alternatively it could be a directed event triggered by the transforming pathway(s). At any rate, many cancer/testis genes are repressed by DNA methylation, but *HORMAD1* and *CT83* are highly specific in their association with basal tumors, so they could be specifically induced in this tumor type, specifically selected for, or both.

Synergy between *HORMAD1* and *CT83*

Using in vitro models of breast cells and gain-of-function tools, we show that the joint expression of *HORMAD1* and *CT83* has transcriptional and phenotypic consequences that are not observed when either gene is expressed in isolation.

In breast cells, joint *HORMAD1/CT83* expression increases the proportion of stem-like cells. This correlates with the induction of a transcriptional signature that is also found in double-positive basal tumors. Therefore, the combined expression of *HORMAD1* and *CT83* is sufficient to increase stem-like properties, and to kick-start a transcriptional program observed in the basal tumors that have the poorest prognosis. This finding suggests that *HORMAD1* and *CT83* are not merely markers but also actors of basal breast tumorigenesis.

We find that neither protein is sufficient to turn on the production of the other, and that neither protein detectably affects the amount or localization of the other. Further work will be necessary to uncover the molecular basis of their synergistic effect.

Limits and perspectives

We note that our analysis has a number of possible limitations. One is that we used pre-existing lists of cancer/testis genes; any gene not detected in these previous publications has not been considered in our work. Another has to do with sensitivity: if certain genes are expressed only in a small number of tumors, then the smoothing we performed in the initial step of our analysis may have made them undetectable. Our sample size was large, with more than 1000 tumors, but certain rare subtypes (such as normal-like tumors, only represented by 40 data points) may benefit from a more focused approach. Also, we focused on one specific type of genes misexpressed in tumors: the cancer/testis genes. However, other tissue-specific genes ectopically expressed in breast tumors can be a rich source of markers and

may be involved in the transformation process. These genes can be easily recovered from our dataset and may deserve further investigations in the future.

In spite of the limitations mentioned above, the current work brings new conceptual insight into the role of cancer/testis genes in breast cancer, showing that reactivation occurs de novo and can have a synergistic effect. In practical terms, as already underlined by other investigators, the genes we have studied represent potential targets for immunotherapy. We show, in addition, that their epigenetic activation seems irreversible, and that they could constitute ideal witnesses of past episodes of epigenetic instability. This may help better understand the role of epigenetic instability in breast tumors, and its mechanistic connection to cellular transformation.

MATERIEL & METHODS

Wet biology

Cell culture

Human mammary cell lines, derived from normal mammary tissue, were obtained from collections developed and generously given by the laboratories of Christophe Ginestier (CRCM) and Raphaël Margueron (Institut Curie). Cancer cell lines (MDA-MB-436, HEK293T) were obtained from ATCC or generously given by the laboratory of Marc-Henri Stern (Institut Curie).

The cell lines were grown using the recommended culture conditions. Cells were incubated in a humidified atmosphere at 37°C under 5% CO₂. All experiments were done with subconfluent cells in the exponential phase of growth.

Cell lines	Medium
HME, HMLE	DMEM:F12 medium supplemented with 10% FBS, 1% penicillin/streptomycin, Non-Essential Amino Acids (LifeTechnology 11140-035) 1%, Insulin Humalog (Lily) 10ug/ml, Hydrocortison (Serb) 0.5 ug/ml, EGF (ThermoFisher PHG0311) 10ng/ml
HEK293T, MDA-MB-436	DMEM medium supplemented with 10% FBS, 1% penicillin/streptomycin

Treatment of cells with 5-aza-dC

Treatment with 5-Aza-dC was performed as described previously (Naciri et al. 2019). Briefly, for dose-response experiments, cells were seeded at a density of $1 \cdot 10^4$ cells in a 6-well tissue culture plate. When cells became firmly adherent to plastic, the medium was replaced with fresh medium containing the appropriate concentration of 5-Aza-dC, every 24h for 2 days (two pulses). For the recovery assay, cells were seeded at a density of XXX in a 100 mm tissue culture plate. When cells became firmly adherent to plastic (T0), the medium was replaced with fresh medium containing 1 uM or 300 nM of 5-Aza-dC for 24h (one pulse). At the end of the treatment, the medium was replaced with fresh culture medium without 5-Aza-dC, and cells were cultured for an additional 2 weeks in subconfluent condition with regular passages. At the end of the treatment and at the appropriate time-points, cells were used for molecular assays. Control cultures were treated under similar experimental conditions in the absence of 5-Aza-dC.

Generation of the *HORMAD1* and/or *CT83* mammary cell lines

The maximal reporter cassette comprised *HORMAD1*-P2A-*CT83*-T2A-Blasti^R (Synthesized by GenScript). The three proteins expressed by the cassette were separated from each other by self-cleaving 2A peptides (P2A, T2A). This cassette was cloned in a lentiviral backbone from ORIGENE (derived from PS100071), under the control of the constitutive CMV promoter. The control plasmid (Blasti^R) and the two other plasmids (*HORMAD1* -T2A-Blasti^R and *CT83*-T2A-Blasti^R) were generated by enzymatic digestion; all the plasmids were grown and prepared individually. The sequences were validated by sequencing. Lentiviruses were generated and used for transduction. Production of lentiviral particles was performed by calcium-phosphate transfection of HEK293T with psPAX2 and pMD2.G plasmids, in a BSL3 tissue culture facility. HME or HMLE cells were seeded into 12-well plates, infected, and selected with blasticidin (5ug/ml) for 15 days.

Western blotting

Cells were harvested and lysed in RIPA buffer (Sigma) with with protease inhibitor cocktail (Thermo Fisher Scientific), sonicated with a series of 30s ON / 30s OFF for 5 min on a Bioruptor (Diagenode), and centrifuged at 16,000 g for 5 min at 4°C. The supernatant was collected and quantified by BCA assay (Thermo Fisher Scientific). Thirty microgram protein

extract per sample was mixed with NuPage 4X LDS Sample Buffer and 10X Sample Reducing Agent (Thermo Fisher Scientific) and denatured at 95°C for 5 min. Samples were resolved on a pre-cast SDS-PAGE 4-12% gradient gel (Thermo Fisher Scientific) with 120V electrophoresis for 90 min and blotted onto a nitrocellulose membrane (Millipore). The membrane was blocked with 5% fat-free milk/PBS at RT for 1 h, then incubated overnight at 4°C with appropriate primary antibodies. After three washes with PBS/0.1% Tween20, the membranes were incubated with the cognate fluorescent secondary antibodies and revealed in the LI-COR Odyssey imaging system. The following antibodies were used in this study: α -HORMAD1 (dilution 1:1000, reference HPA037850), α -CT83 (dilution 1:1000, reference HPA004773), α -Tubulin (dilution 1:10 000, reference Abcam ab7291).

Quantitative Real-time PCR

RNA extraction was done using Tri reagent according to the manufacturer's recommendations. One microgram of total RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) and Oligo dT primers (Promega). qPCR was performed using Power SYBR Green (Applied Biosystems) on a Viia 7 Real-Time PCR System (Life Tech). *TBP* and *PGK1* genes were used for normalization of expression values. Primer sequences are available in Supplementary Table S4.

Immunohistochemistry

Demande à Didier les conditions XP.

Breast cancer cohort from Institut Curie, 99 samples, each sample was marked with the two antibodies (HORMAD1 and CT83)

Immunofluorescence

Cells were grown on glass coverslips, fixed with 4% paraformaldehyde (PFA) for 10 minutes, and then permeabilized with 0.5% Triton. The glass coverslips were then blocked with 1% bovine serum albumin in phosphate buffer saline for 1 hour, before applying primary antibody for 1 hour. After this incubation, secondary antibody was applied for 45 minutes, before washed and applied Hoechst stain (1:20,000; Sigma #33258). The following antibodies were used: Anti-HORMAD1 (HPA037850; 1/3000), Anti-CT83 (AMAB91318; 1/200); Donkey anti-Rabbit Alexa Fluor 488 (1/2000), Donkey anti-Mouse Alexa Fluor 594 (1/200).

Flow Cytometry

Freshly dissociated cells were stained with APC-conjugated EpCAM (dilution 1:100, Miltenyi clone HEA-125) and PerCP-Cy5.5-conjugated CD49f (dilution 1:10, BD clone GoH3); or APC-conjugated CD44 (dilution 1:10, BD clone G44-CD26) and PerCP-Cy5.5-conjugated CD24 (dilution 1:10, BD clone ML5); with live/dead Violet (dilution 1:1000, ThermoFisher) for cell viability, in HBSS (Gibco) with 2% FBS and incubated at room-temperature for 20 min, followed by washing in HBSS with 2% FBS and re-suspended in HBSS/FBS 2%. Analysis was performed by using a CyAn (Beckman Coulter) flow cytometer. Thresholds on fluorescence signal intensity (subtracting background fluorescence from the appropriate isotype control antibodies) were used to determine the proportion of cell populations. Data were analyzed with FlowJo software.

Wound healing assay

Each HMLE cell lines were plated in 12-well dishes and grown to confluence. A scratch was performed in the middle of the well using a P10 plastic tip and image acquisition was performed every hour using the Incucyte live cell system. Images were analyzed using ImageJ software and the following parameters: Analyze Particles (size = 0-Infinity), after the following protocol: Process => Sharpen & Find Edges, Threshold 132-255, Process => Binary => convert to mask.

Bioinformatics

Public data sets used in this study

We used previously published gene lists to define testis-specific genes, tumor suppressor genes and oncogenes. We also used multiple public datasets involving both normal and tumor tissues to evaluate C/T gene expression. Detailed information of these databases was listed in the Supplementary Table 5.

Development of the Cancer-Gene Markers Detection pipeline

Briefly, we computed the Kernel's density estimation for each gene expression pattern in healthy mammary gland and in breast cancer cohorts, respectively. We then analyzed density profiles variations using the derivative of the density functions, and classify genes as unimodal or multimodal in normal mammary tissues and breast cancer samples. For each gene, we calculated the mean expression values in normal and cancers samples. We classify genes according to these parameters, as described in figure S1A-C. All the detailed scripts are available on GitHub (https://github.com/MartheLaisne/CTA_BreastCancers).

Identification of genes with abnormal breast cancer expression pattern using transcriptomic TCGA analysis

TCGA gene count datasets for breast normal and cancer samples were downloaded using TCGAblinks (Colaprico et al., v2.10.5). Expression were normalized with DESeq2 (Love MI, Huber W, Anders S, v.1.22.2). Abnormally expressed genes were defined as any expression value greater than the mean expression + 3 standard-deviations in normal mammary tissues. All the detailed scripts are available on GitHub.

Validation of the Testis-specific expression pattern for the selected 139 C/T genes

Expression values for GTEx (Carithers LJ et al., 2015) dataset was obtained directly from the project webpage as TPM values, and the median expression values by tissue were calculated. We extracted expression values for the 139 selected TS genes, and we performed an unsupervised clustering (Euclidean distance and complete method) of the genes and the samples based on these values. Detailed script is on GitHub.

Analyze of the INVADE dataset

Briefly, raw counts were normalized using DESeq2 (Love MI, Huber W, Anders S, v.1.22.2). because there are no normal tissues in this dataset, another strategy was used to defined the threshold for abnormal C/T gene activation: we used the bimodality of the expression values distribution to define a background level. Any expression value below this threshold was considered as noise, and the gene as repressed. The top 20 CT genes based on random forest analyzes were used to performed an unsupervised hierarchical clustering (binary distance and Ward.D2 method) of the 55 tumors samples. Detailed script is on GitHub.

Survival and drug-response analysis

For recidive-free survival (RFS), data were download from <https://kmplot.com/analysis/> (n=4934), using the indicated parameters for sample selection. Data were then analyzed using custom R script and survminer and survival R packages. For anthracyclin-response analysis, data were dowload from <http://www.rocplot.org/> using the indicated parameters for sample selection, and analyzed using standard R functions. ROC curves were generated using ROCit R package.

Analyze of normal mammary breast microarray

Data were download at <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4145>. The raw CEL data were normalized using the following packages: affy (v1.60.0), ArrayExpress (v1.42.0) for annotation and data importation; oligo (v1.45.0), arrayQualityMetric (v3.38.0) for quality control and pre-processing; limma (v3.38.3) for analysis and statistics.

scRNAseq of normal mammary breast cells

Briefly, data were download (GSE113197) and analyze using Seurat (v3.1.4) package. For the normalization, we keep unexpressed genes because we are interested in C/T genes, which are expected to not be expressed in healthy mammary cells. We filtered cells to keep only cell with at least 500 genes detected, but no more than 6000, and less than 10% of mitochondrial gene expressed. UMAP was performed using the 10 first components of the PCA. Cell identities were assigned based on the expression of lineage markers (source code is at: https://github.com/Michorlab/tnbc_scrnaseq/blob/master/code/funcs_markers.R) . Detailed script is on GitHub.

scRNAseq of triple-negative breast tumors

FASTQ read pairs were aligned to the human reference genome (build gencode v29) using STAR (v2.7.5c) and default single-pass parameters. Uniquely aligned reads were kept for downstream analysis using Samtools view (v1.10) and parameters: -q 10 -b -o, and counted with htseq (--stranded=yes --type=exon). Data were analyzed using Seurat (v3.1.4). As for Healthy mammary scRNAseq analyze, we identified low quality cells by (i) few expressed genes, (ii) abnormally high number of expressed genes and (iii) high mitochondrial gene expression. Cell identities were determined using the same procedure than for the healthy mammary scRNAseq data. We also used Lehman signature to assigned each cancer cell to a lehman subtype, as described in the original publication (code source: https://github.com/Michorlab/tnbc_scrnaseq)

Differential Gene Expression Analysis in TCGA basal-like samples

HORMAD1- and CT83-positive tumors were identified based on normalized RNAseq (FPKM-UQ) data downloaded from TCGA (2020 accession). Briefly, we defined a threshold for positive HORMAD1 and CT83 expression based on the expression level detected in non-tumor breast samples (NT) as follow:

$$Thr_{CT} = Mean_{CT}(NT) + 2 * SD_{CT}(NT)$$

We classified tumors in 4 different groups based on their expression levels of both HORMAD1 and CT83. Then, we download HTseq-counts data for basal-like breast tumors only and we performed a differential expression analysis using the R package *DESeq2*, with the HORMAD1 & CT83 label as factor of interest. Differentially expressed genes were defined with p-adjusted < 0.05 and absolute value for the fold-change > 1.5.

Differential Peaks Intensity Analysis in TCGA basal-like samples

Both raw counts ATAC-seq data and gene expression data from TCGA were accessed (2020 accession) through either the Genomic Data Commons (GDC) using the GDC Data Transfer Tool Client or the data transfer tool TCGAbiolinks (Colaprico 2016). Individual patient files were assembled using in-house scripts in an R computing environment. Preprocessing consisted of patient and gene matching between data types, log transformation of gene expression data, and classification of the ATAC-seq samples regarding to their HORMAD1 / CT83 expression status, defined in the previous section. For differential analysis, we basal-like tumors from ATAC-seq datas (n=30). Differential peak intensities were found using *DESeq2*. Differentially open regions were defined with p-adjusted < 0.01 and absolute value for the fold-change > 2.

CpG promoter classes identification

Promoters were according to the hg38 version of the human genome, as described in the original article (Weber et al. 2007). Briefly, promoters were classified in three categories to distinguish strong CpG islands, weak CpG islands and sequences with no local enrichment of CpGs. We determined the GC content and the ratio of observed versus expected CpG dinucleotides in sliding 500-bp windows with 5-bp offset. The CpG ratio was calculated using the following

formula: (number of CpGs × number of bp) / (number of Cs × number of Gs). The three categories of promoters were determined as follows: HCPs (high-CpG promoters) contain a 500-bp area with CpG ratio above 0.75 and GC content above 55%; LCPs (low-CpG promoters) do not contain a 500-bp area with a CpG ratio above 0.48; and ICPs (intermediate CpG promoters) are neither HCPs nor LCPs.

Correlation DNA methylation data and expression data for TCGA samples

Both DNA methylation data, Copy Number Variations (CNV) data and gene expression data from TCGA were accessed (2020 accession) through either the Genomic Data Commons (GDC) using the GDC Data Transfer Tool Client or the data transfer tool TCGAAbiolinks (Colaprico 2016). Individual patient files were assembled using in-house scripts in an R computing environment. Preprocessing consisted of patient and gene matching between data types and log transformation of gene expression data. The methylation data in this study were acquired by the Illumina 450K array, which interrogates more than 450 000 methylation sites on the Illumina chip. The data for this study contained information of 485 578 CpG sites. The CNV data were acquired by the Affymetrix SNP 6.0 array numeric CNV values were derived from GISTIC2.

Correlation analysis was performed using Pearson's correlation. The correlation was performed between methylation beta values (respectively between CNV values) and log-base-2-transformed gene expression data with a *p*-value threshold of 0.05. All statistical tests used standard R functions.

Correlation adjacent genes TCGA

Correlation analysis was performed using Pearson's correlation. The correlation was performed between the two log₂ normalized adjacent genes expression values. All statistical tests used standard R functions.

RNA-sequencing: Library preparation for transcriptome sequencing

A total amount of 1 µg total RNA per sample was used as input material for the RNA sample preparations. RNA samples were spiked with ERCC RNA Spike-In Mix (Thermo Fisher Scientific). Sequencing libraries were generated using NEB-Next Ultra™ RNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were replaced by dUTP. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially 250-300 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter). Then 3 µl USER Enzyme (NEB) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. At last, products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

RNA-sequencing: read alignment

FASTQ reads were trimmed using Trimmomatic (Bolger et al., 2014) (v0.39) and parameters: ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:36. Read pairs that survived trimming were aligned to the human reference genome (build hg38) using STAR (Dobin et al., 2013) (v2.7.5c) and default single-pass parameters. PCR duplicate read alignments were flagged using Picard-tools (2019) MarkDuplicates (v2.23.4). Uniquely aligned, non-PCR-duplicate reads were kept for downstream analysis using Samtools (Li et al., 2009) view (v1.10) and parameters: -q 255 -F 1540. Gene expression values were calculated over the hg38 NCBI RefSeq Genes annotation using VisRseq (Younesy et al., 2015) (v0.9.12) and normalized per million aligned reads per transcript length in kilobases (RPKM). Bigwig files were

generated using deeptools (Ramírez et al., 2016) bamCoverage (v3.3.0) using counts per million (CPM) normalization and visualized in IGV (Thorvaldsdottir et al., 2013) (v2.8.9).

RNA-seq: Differential expression, PCA plots, and heatmaps

All the analysis and figures were generated using custom scripts and R version 3.5.2. Scripts are available on Github (https://github.com/MartheLaisne/CTA_BreastCancers/).

Gene set enrichment analysis (GSEA)

Gene set enrichment analysis was performed using GSEA (Subramanian et al., 2005; Mootha et al., 2004) (v4.1.0), msigdb and fgsea package and default parameters (1000 permutations, permutation type = gene_set. Selected significant terms from Hallmark gene sets (n=50), KEGG gene set (n=186), GO biological functions (n =1001) and Curated Breast Pathways (n=169) were displayed. Curated breast gene set is available in Supplementary Table 3.

ACKNOWLEDGEMENTS

We are grateful to the following colleagues for helpful discussions: Saadi Khochbin, Fatima Mechta-Grigoriou, Marc-Henri Stern, Raphael Margueron, Laia Richart, Céline Vallot, Josh Waterfall, Julie Cocquet, Valérie Borde, Cathy Jackson, Jafar Sharif, Julien Sage, Bernard de Massy.

Plateformes : Cytométrie IJM, Epigénétique EDC, Microscopie IJM/EDC

Funding: the team of PAD gratefully acknowledges funding by Fondation pour la Recherche Médicale and by Fondation ARC (PGA1 RF20180206807). ML was the recipient of a 4th-year doctoral fellowship from Ligue Nationale contre le Cancer.

REFERENCES

- Adélaïde, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D., et al. (2007). Integrated profiling of basal and luminal breast cancers. *Cancer Res* 67, 11565–11575.
- Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C.C., Mundstein, A.S., Cohen, T., Chen, Y.-T., Chua, R., Gurung, S., Gnjatic, S., et al. (2009). CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37, D816–D819.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795.
- Chen, B., Tang, H., Chen, X., Zhang, G., Wang, Y., Xie, X., and Liao, N. (2018). Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag Res* 11, 179–190.
- Chen, C., Gao, D., Huo, J., Qu, R., Guo, Y., Hu, X., and Luo, L. (2021a). Multiomics analysis reveals CT83 is the most specific gene for triple negative breast cancer and its hypomethylation is oncogenic in breast cancer. *Sci Rep* 11, 12172.
- Chen, C., Gao, D., Huo, J., Qu, R., Guo, Y., Hu, X., and Luo, L. (2021b). Multiomics analysis reveals CT83 is the most specific gene for triple negative breast cancer and its hypomethylation is oncogenic in breast cancer. *Sci Rep* 11, 12172.
- Chen, Z., Zuo, X., Pu, L., Zhang, Y., Han, G., Zhang, L., Wu, Z., You, W., Qin, J., Dai, X., et al. (2019). Hypomethylation-mediated activation of cancer/testis antigen KK-LC-1 facilitates hepatocellular carcinoma progression through activating the Notch1/Hes1 signalling. *Cell Prolif.* 52, e12581.
- Chung, W., Eum, H.H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., Ryu, H.S., Kim, S., Lee, J.E., Park, Y.H., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 8, 15081.
- Date, S., Nozawa, O., Inoue, H., Hidema, S., and Nishimori, K. (2012). Impairment of pachytene spermatogenesis in *Dmrt7* deficient mice, possibly causing meiotic arrest. *Biosci Biotechnol Biochem* 76, 1621–1626.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* 15, 50–65.
- Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* 18, 197–218.
- Gibbs, Z.A., and Whitehurst, A.W. (2018). Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors. *Trends Cancer* 4, 701–712.
- Greenbaum, M.P., Yan, W., Wu, M.-H., Lin, Y.-N., Agno, J.E., Sharma, M., Braun, R.E., Rajkovic, A., and Matzuk, M.M. (2006). *TEX14* is essential for intercellular bridges and fertility in male mice. *Proc Natl Acad Sci U S A* 103, 4982–4987.
- Györfy, B. (2021). Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Computational and Structural Biotechnology Journal* 19, 4101–4109.
- Holm, K., Staaf, J., Lauss, M., Aine, M., Lindgren, D., Bendahl, P.-O., Vallon-Christersson, J., Barkardottir, R.B., Höglund, M., Borg, Å., et al. (2016). An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. *Breast Cancer Res* 18, 27.
- Karlin, K.L., Mondal, G., Hartman, J.K., Tyagi, S., Kurley, S.J., Bland, C.S., Hsu, T.Y.T., Renwick, A., Fang, J.E., Migliaccio, I., et al. (2014). The oncogenic STP axis promotes triple-negative breast cancer via degradation of the REST tumor suppressor. *Cell Rep* 9, 1318–1332.
- Kaufmann, J., Wentzensen, N., Brinker, T.J., and Grabe, N. (2019). Large-scale in-silico identification of a tumor-specific antigen pool for targeted immunotherapy in triple-negative breast cancer. *Oncotarget* 10, 2515–2529.
- Kondo, Y., Fukuyama, T., Yamamura, R., Futawatari, N., Ichiki, Y., Tanaka, Y., Nishi, Y., Takahashi, Y., Yamazaki, H., Kobayashi, N., et al. (2018). Detection of KK-LC-1 Protein, a Cancer/Testis Antigen, in Patients with Breast Cancer. *Anti-cancer Res* 38, 5923–5928.

- Kumar, R., Ghyselinck, N., Ishiguro, K., Watanabe, Y., Kouznetsova, A., Höög, C., Strong, E., Schimenti, J., Daniel, K., Toth, A., et al. (2015). MEI4 – a central player in the regulation of meiotic DNA double-strand break formation in the mouse. *J. Cell. Sci.* 128, 1800–1811.
- Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., and Pietenpol, J.A. (2016). Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One* 11, e0157368.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liu, Y., DeBoer, K., de Kretser, D.M., O'Donnell, L., O'Connor, A.E., Merriner, D.J., Okuda, H., Whittle, B., Jans, D.A., Efthymiadis, A., et al. (2015). LRGUK-1 Is Required for Basal Body and Manchette Function during Spermatogenesis and Male Fertility. *PLoS Genet* 11, e1005090.
- Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R.S., and Cusack, S. (2012). The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA* 18, 2056–2072.
- Mischo, A., Kubuschok, B., Ertan, K., Preuss, K.-D., Romeike, B., Regitz, E., Schormann, C., Bruijn, D. de, Wadle, A., Neumann, F., et al. (2006). Prospective study on the expression of cancer testis genes and antibody responses in 100 consecutive patients with primary breast cancer. *International Journal of Cancer* 118, 696–703.
- Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 7, 403–417.
- Morel, A.-P., Ginestier, C., Pommier, R.M., Cabaud, O., Ruiz, E., Wicinski, J., Devouassoux-Shisheboran, M., Combaret, V., Finetti, P., Chassot, C., et al. (2017). A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med* 23, 568–578.
- Muñoz, F., Martínez-Jiménez, F., Pich, O., Gonzalez-Perez, A., and Lopez-Bigas, N. (2021). In silico saturation mutagenesis of cancer genes. *Nature*.
- Naciri, I., Laisné, M., Ferry, L., Bourmaud, M., Gupta, N., Di Carlo, S., Huna, A., Martin, N., Peduto, L., Bernard, D., et al. (2019). Genetic screens reveal mechanisms for the transcriptional regulation of tissue-specific genes in normal cells and tumors. *Nucleic Acids Res* 47, 3407–3421.
- Nichols, B.A., Oswald, N.W., McMillan, E.A., McGlynn, K., Yan, J., Kim, M.S., Saha, J., Mallipeddi, P.L., LaDuke, S.A., Vilalobos, P.A., et al. (2018). HORMAD1 Is a Negative Prognostic Indicator in Lung Adenocarcinoma and Specifies Resistance to Oxidative and Genotoxic Stress. *Cancer Res.* 78, 6196–6208.
- Paret, C., Simon, P., Vormbrock, K., Bender, C., Kölsch, A., Breitkreuz, A., Yildiz, Ö., Omokoko, T., Hubich-Rau, S., Hartmann, C., et al. (2015). CXorf61 is a target for T cell based immunotherapy of triple-negative breast cancer. *Oncotarget* 6, 25356–25367.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160–W165.
- Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.-Y., Lantuejoul, S., Hainaut, P., et al. (2013a). Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* 5, 186ra66.
- Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.-Y., Lantuejoul, S., Hainaut, P., et al. (2013b). Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci Transl Med* 5, 186ra66.
- Shin, Y.-H., Choi, Y., Erdin, S.U., Yatsenko, S.A., Kloc, M., Yang, F., Wang, P.J., Meistrich, M.L., and Rajkovic, A. (2010). Hormad1 Mutation Disrupts Synaptonemal Complex Formation, Recombination, and Chromosome Segregation in Mammalian Meiosis. *PLOS Genetics* 6, e1001190.
- Sohni, A., Tan, K., Song, H.-W., Burow, D., de Rooij, D.G., Laurent, L., Hsieh, T.-C., Rabah, R., Hammoud, S.S., Vicini, E., et al. (2019). The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Rep* 26, 1501-1517.e4.
- Varley, K.E., Gertz, J., Roberts, B.S., Davis, N.S., Bowling, K.M., Kirby, M.K., Nesmith, A.S., Oliver, P.G., Grizzle, W.E., Forero, A., et al. (2014). Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res Treat* 146, 287–297.
- Wang, C., Gu, Y., Zhang, K., Xie, K., Zhu, M., Dai, N., Jiang, Y., Guo, X., Liu, M., Dai, J., et al. (2016). Systematic identifi-

cation of genes with a cancer-testis expression pattern in 19 cancer types. *Nat Commun* 7, 10499.

Wang, J., Rousseaux, S., and Khochbin, S. (2014). Sustaining cancer through addictive ectopic gene activation. *Curr Opin Oncol* 26, 73–77.

Wang, X., Tan, Y., Cao, X., Kim, J.A., Chen, T., Hu, Y., Wexler, M., and Wang, X. (2018). Epigenetic activation of *HORMAD1* in basal-like breast cancer: role in Rucaparib sensitivity. *Oncotarget* 9, 30115–30127.

Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., et al. (2015a). Genomic Complexity Profiling Reveals That *HORMAD1* Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. *Cancer Discov* 5, 488–505.

Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., et al. (2015b). Genomic complexity profiling reveals that *HORMAD1* overexpression contributes to homologous recombination deficiency in triple-negative breast cancers. *Cancer Discov* 5, 488–505.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39, 457–466.

Whitehurst, A.W. (2014). Cause and consequence of cancer/testis antigen activation in cancer. *Annu Rev Pharmacol Toxicol* 54, 251–272.

Zhong, G., Lou, W., Shen, Q., Yu, K., and Zheng, Y. (2020). Identification of key genes as potential biomarkers for triple-negative breast cancer using integrating genomics analysis. *Mol Med Rep* 21, 557–566.

TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data - PubMed.

Unified single-cell analysis of testis gene regulation and pathology in five mouse strains - PubMed.

VisRseq: R-based visual framework for analysis of sequencing data - PubMed.

FIGURE LEGENDS

Figure 1: A custom bioinformatic screen identifies 139 Cancer/Testis genes abnormally expressed in breast tumors

- A. Schematic description of the bioinformatic pipeline. We depict the expression profile of a gene that passed the screen: it has a unimodal, zero-centered profile in normal tissue, and a multimodal profile in breast tumors.
- B. Chow-Ruskey diagram showing the intersection between previously published C/T gene lists and the C/T genes that were selected for our study.

Figure S1: Optimizing parameters for the bioinformatic screen, further uses and validations

- A. Outputs of the screen for different smoothing parameters (Bandwidth). Previously known breast cancer markers (*ESR1*, *PGR*, *ERBB2*) were used as positive controls, and housekeeping genes (*ACTB*, *GAPDH*, *TUBA1A*) were used as negative control. A red minus sign means the gene was not detected as aberrantly expressed in tumors, a green plus sign means that it was. The total number of atypically expressed genes for each bandwidth value is shown.
- B. Classification of all genes according to our parameters: we were interested in genes with a homogeneous expression in NB (*ie.* Unimodal profile in NB). Then, these genes can be subsequently divided according to their expression pattern in breast tumors: two situations were of specific interest: genes that are homogeneously expressed in breast tumors too (panel C), and genes that are overexpressed or repressed in a subset of breast tumors (panel D).
- C. Refinement of the characterization of homogeneously expressed genes in NB and in breast tumors, respectively: when means were significantly different in NB and in Tum, these genes could be used as tumor markers. Some of such genes are known overexpressed oncogenes or repressed tumor suppressor genes; a significant part of them (1362 genes) are unknown but could play a role in breast tumor development
- D. Refinement of the characterization for tumor-specific variables genes: approximately 70% of them are repressed in NB and abnormally activated in breast tumors; amongst these genes there are known tissue-specific genes (including testis-specific genes). The remaining 30% are overexpressed or repressed genes in some breast tumors, including known subtype-specific oncogenes like *ESR1*, and others genes that could be used as marker of specific tumor subgroups.
- E. Heatmap showing the mean expression values (Z-score) for the 139 selected C/T genes in various human adult tissues, based on RNA-seq data from GTEx.

Figure 2: The activation of specific C/T genes is predictive of tumor subtype, occurs early during tumorigenesis, and is associated with prognosis

- A. Multidimensional analysis of TCGA breast tumor and healthy samples based on expression of the 139 selected C/T genes. Each dot is a sample, the color code corresponds to the tumor subtype by PAM50 molecular classification. Left: Principal Component Analysis, dot sizes are proportional to the quality of representation in PC1/PC2 space. The C/T genes best correlated to PC1/PC2 are represented. Right: Uniform Manifold Approximation and Projection (UMAP).
- B. Confusion matrix for breast tumor samples in the validation cohort (25% of the samples, randomly selected from the TCGA breast tumors), using the the best Random Forest model. This model was established after a 500-tree training on the discovery cohort (75%), based on the expression level of the 139 C/T genes.
- C. The top 15 most important variables in the best Random Forest model for PAM50 subtype prediction. The color of the gene name indicates the tumor type most associated.
- D. Expression levels for 6 subtype-specific C/T genes in the breast TCGA cohort, according to PAM50 tumor subtype.
- E. Hierarchical clustering of early and late breast tumor samples, based on expression of the top 6 C/T genes described earlier. A C/T gene is depicted as activated (black box) if its expression value is above the background expression threshold.
- F. Relapse-free survival curves for ER+ Her2- breast cancer patients according to LRGUK expression, for Luminal A tumors (left), and for Luminal B tumors (right).
- G. Left: Expression value for the luminal-specific C/T gene LRGUK in luminal B tumors, according to the clinical

evaluation of tumor response to chemotherapy. Right: ROC curve evaluating the potential of LRGUK as a predictive biomarker of anthracyclin chemotherapy response of ER+ Her2- Luminal B tumors.

- H. Relapse-free survival curve for ER-PR-Her2- Basal-like breast cancer patients, as a function of HORMAD1 expression alone, CT83 expression alone, or combined expression of the two C/T genes.
- I. Left: Combined expression value for the two basal-specific C/T genes HORMAD1 and CT83 in basal-like tumors, according to the clinical evaluation of tumor response to chemotherapy. Right: ROC curve evaluating the potential of HORMAD1 and CT83 combined expression as a predictive biomarker of anthracyclin chemotherapy response of ER- PR- Her2- Basal-like tumors.

Figure S2 Examination of marker expression in tumors classified by IHC and in tumor cell lines. Expression in early tumors and association with survival.

- A. Multidimensional analysis of TCGA breast tumor and healthy samples based on the 139 selected C/T gene expression. Each dot is a sample, color code corresponds to immunohistochemistry (IHC) classification (based on ER/PR/HER2 expression). Left: Principal Component Analysis, dot sizes are proportional to the quality of representation in PC1/PC2 space. The best correlated C/T genes to PC1/PC2 are represented. Right: Uniform Manifold Approximation and Projection
- B. Confusion matrix for breast tumor samples in the validation cohort (randomly selected 25% samples from the TCGA breast tumors) of the IHC tumor subtypes prediction obtained with the best Random Forest model. This model was established after a 500 trees training on the discovery cohort (the remaining 75%), based on the expression level of the 139 C/T
- C. Top 15 most important variables in the best Random Forest model for IHC tumor subtype prediction.
- D. Expression levels for the 2 basal-specific C/T genes in the breast TCGA cohort, according to IHC tumor subtype
- E. Expression levels for 6 subtype-specific C/T genes in breast cancer cell lines from the Cancer Cell Line Encyclopedia, according to PAM50 tumor subtype. Some commonly used cell lines are highlighted.
- F. Hierarchical clustering of early and late breast tumor samples, based on the the 39 C/T genes that are expressed in at least one tumor. A C/T gene is depicted as activated if its expression value is above the background expression threshold.
- G. Relapse-free survival curves for Her2-positive breast cancer patients, according to DMRTC2 expression (left), or TDRD1 expression (right).
- H. Immune infiltration of Her2-positive breast tumors that express (ON) or do not express (OFF) DMRTC2, inferred from whole tumor RNA-seq data using MCPcounter. Fold-Change were computed against Normal Breast (NB). Right: Expression level of the immune suppressive factor *FOXP3* in the same tumors. P-value < 0.01: ** ; P-value < 0.001: ***
- I. Same as panel H, but for basal-like tumors that either express (ON) or do not express (OFF) HORMAD1 and CT83.

Figure 3: HORMAD1 and CT83 are expressed specifically by cancer cells, however scRNA-seq reveals rare HORMAD1+ / CT83+ luminal progenitor cells in healthy mammary gland

- A. Top 15 most important variables in the best Random Forest model applied to an independent cohort of breast tumors .
- B. Expression of HORMAD1 and CT83 in the indicated sample types of the Varley/Myers cohort (GSE58135)
- C. Co-expression of *HORMAD1* and *CT83* based on RNA-seq analysis (log2 FPKM-UQ) in basal-like breast tumor samples (n=194) from the TCGA. Threshold for positive or negative expression are calculated based on the corresponding gene expression profile in tumors at the second inflexion point of the representative curve. The number of tumors belonging to each category is shown.
- D. Immunohistochemistry of HORMAD1 (left) or CT83 (right) in testis (1, 2) and triple-negative breast cancer sample (3, 4).
- E. UMAP representation of a scRNA-seq study on 6 triple-negative breast tumors (GSE75688). Each dot is either a tumor cell or a cell from the tumor microenvironment. From left to right: cell types which were determined based on the expression of specific marker genes; *HORMAD1* normalized expression level; *CT83* normalized expression level.

- F. Schematic representation of the mammary cell hierarchy in healthy adult mammary gland.
- G. *HORMAD1* and *CT83* expression in sorted healthy mammary cells. The red dotted line represents the threshold for gene expression detection.
- H. UMAP representation of a scRNA-seq study on 4 healthy mammary glands (GSE113197), after an enrichment in epithelial cell by FACS. From left to right: cell types which were determined based on the expression of specific marker genes; *HORMAD1* normalized expression level; *CT83* normalized expression level.

Figure S3: Characteristics of *HORMAD1/CT83*-positive basal tumors and cell lines, validation by IHC

- A. Characteristics of basal-like breast tumors from the TCGA according to their activation status of *HORMAD1* and *CT83*.
- B. Links between *HORMAD1* and *CT83* expression and Lehman's basal tumor subgroups
- C. Co-expression of *HORMAD1* and *CT83* based on RNA-seq analysis (log2 Normalized expression) in basal-like breast cancer cell lines (n= 22) from the CCLE. Thresholds were calculated as in Fig. 3A.
- D. Left: Co-expression of *HORMAD1* and *CT83* based on immunohistochemistry analysis on Breast Cancer Cohort (n= 88) from the Curie Institute. Thresholds were calculated as in Fig. 3A. Tumors were classified according to ER/PR/HER2 and Ki67 expression. Right: proportion of breast tumors in each *HORMAD1* & *CT83* expression categories.
- E. UMAP representation of a scRNA-seq study on 4 healthy mammary glands (GSE113197), after an enrichment in epithelial cell by FACS. *MSRB3* or *ESR1* expression marks the expected populations.

Figure 4: *HORMAD1* & *CT83* expressions are epigenetically regulated, with an essential contribution of DNA methylation

- A. Correlation between *HORMAD1* and *CT83* expression and mean DNA methylation of their promoters (TSS +/- 200bp), according to the copy number variation of their genomic loci.
- B. Correlation between *HORMAD1* or *CT83* expression, and the expression of their neighboring genes. *ERBB2* is a positive control. The color code corresponds to PAM50 tumor subtypes
- C. IGV representation of *HORMAD1* and *CT83* genomic loci, with CpG density promoter classification according to the Weber/Schübeler criteria (PMID: **17334365**). ATAC-seq data are from representative basal-like tumors (TCGA cohort). Differentially accessible regions (DAR) between these two groups of basal tumors were identified.
- D. Inverse correlation between *HORMAD1* and *CT83* expression and the mean DNA methylation of their promoters (TSS +/- 200bp). Each dot represents a tumor, and the color intensity indicates Copy Number Variation.
- E. RTqPCR analysis of *HORMAD* and *CT83* expression in non-tumorigenic human mammary cell lines, in control condition or following a 48 hours 5-Aza-dC treatment at various concentrations.
- F. Western Blot of *HORMAD1* and *CT83* expression in non-tumorigenic human mammary cell lines, in control condition or following a 48 hours 5-Aza-dC treatment at 0.3 μ M.
- G. RT-qPCR analysis of *HORMAD* and *CT83* expression at various time points, in the same cell line, after an initial perturbation with 0.3 or 1 μ M 5-Aza-dC followed by a recovery period in drug-free medium.

Figure S4: Epigenetic landscapes of the *HORMAD1* and *CT83* genes

- A. IGV representation of transcriptomic and histone modification landscapes at *HORMAD1* and *CT83* loci in healthy Testis and Breast samples (ENCODE).
- B. DNA methylation levels on the promoter of *HORMAD1* or *CT83*, in normal human breast and sperm, from 450K array values.
- C. Total accessibility scores for *HORMAD1*- and *CT83*-negative or positive basal-like TCGA tumors, calculated based on ATAC-seq data.
- D. IGV representation of histone modifications landscapes at *HORMAD1* and *CT83* loci in breast cell lines: MCF7 cells do not express *HORMAD1* or *CT83*, whereas MDA-MB436 cells express both.

Figure 5: **HORMAD1 & CT83 act synergistically to promote aggressive features in non-transformed mammary cells**

- A. Cellular model and lentiviral vectors used to express HORMAD1 and/or CT83. P2A and T2A are self-cleaving peptides. BsR: Blastidicin-resistance gene.
- B. *HORMAD1*, *CT83* and the reporter *BsR* mRNA expression, assessed by RT-qPCR, in HMLE-derived cell lines. Data are represented as means +/-sd (n=3 independent experiments).
- C. Western Blot analysis of HORMAD1 and CT83 expression in the indicated HMLE-derived cell lines.
- D. Immunofluorescence staining of HORMAD1 and CT83 in HMLE-derived cell lines.
- E. FACS analysis of EpCAM and CD49f cell-surface markers in HMLE-derived cell lines. Mature luminal cells (mL, EpCAM⁺CD49f⁺), luminal progenitor cells (Lp, EpCAM⁺CD49f⁺), mammary stem cells (MaSC, EpCAM⁺CD49f^{low}) One experiment representative of 3 independent experiments is shown. Bottom: Summary of 3 independent experiments (Mean +/- SD)
- F. FACS analysis of CD24 and CD44 cell-surface markers in HMLE-derived cell lines. Mature cells (mat, CD44⁺CD24^{high}), intermediate cells (int, CD44⁺CD24^{low}), Stem cells (SC, CD44⁺CD24^{low}) One experiment representative of 3 independent experiments is shown. Bottom: Summary of 3 independent experiments (Mean +/- SD)

Figure S5: **Additional experiments on HORMAD1 and CT83 in cultured breast cells**

- A. Immunofluorescence staining of E-CADHERIN and endogenous signal from HORMAD1-GFP reporter construct in HMLE-derived cell lines. Cells were sorted 48h after the infection based on the detection of the GFP signal, and immediately seeded for staining.
- B. Immunofluorescence staining of E-CADHERIN and endogenous signal from CT83-GFP reporter construct in HMLE-derived cell lines. Cells were sorted 48h after the infection based on the detection of the GFP signal, and immediately seeded for staining.
- C. Growth curves of HMLE-derived cell lines, measured by indirect detection of metabolically active cells using MTS assay (left, means +/- sd, n=3 independent experiments)
- D. Schematic representation of the non-tumorigenic mammary cell lines (HME) generated for this study
- E. *HORMAD1*, *CT83* and the reporter *BLASTICIDIN* mRNA expression, assessed by RT-qPCR, in HME-derived cell lines. Data are represented as means +/-sd (n=3 independent experiments).
- F. FACS analysis of EpCAM and CD49f cell-surface markers in HME-derived cell lines. Mature luminal cells (mL, EpCAM⁺CD49f⁺), luminal progenitor cells (Lp, EpCAM⁺CD49f⁺), mammary stem cells (MaSC, EpCAM⁺CD49f^{low}).

Figure 6: **HORMAD1 & CT83-induced phenotype in HMLE cell line is also found in double-positive basal breast tumors**

- A. Volcano plot of the distribution of differentially expressed genes (p-value adjusted < 0.01) in HMLE cells expressing HORMAD1, CT83, or both, relative to control cells. Certain genes of particular interest are indicated.
- B. Barplot displaying the top 6 pathways differentially activated in HORMAD1 and CT83-positive HMLE-derived cells compared to the control condition, from MSigDB c2_curated Breast, c7_Hallmark, c5_GO and c2_KEGG annotations. x-axis corresponds to -log₁₀ adjusted q-values.
- C. GSEA analysis comparing the «HORMAD1+CT83» signature detected in HMLE cells (88 genes, see panel A), to the ranked transcriptome of the double-positive cell lines in the CCLE dataset, or of double-positive tumors in the TCGA dataset.

Figure S6: **Validation of the “HORMAD1+CT83” signature, additional analyses in tumors and cell lines**

- A. qRT-PCR experiments in the indicated cell lines to validate the RNA-seq results of Figure 6A.
- B. qRT-PCR shows a decrease of CDH1 expression upon joint HORMAD1/CT83 expression in HME cells.
- C. Volcano plot of the distribution of all differentially expressed genes (p-value adjusted < 0.01) in HORMAD1- and CT83-positive basal breast cancer cell lines (n=8), compared to negative basal breast cancer cell line (n=7) from the CCLE, mapping the 200 upregulated genes (red) and the 240 downregulated genes (blue).

- D. Volcano plot of the distribution of all differentially expressed genes (p-value adjusted < 0.01) in *HORMAD1*- and CT83-positive basal tumors (n=111), compared to negative basal breast tumors (n=16) from the TCGA, mapping the 157 upregulated genes (red) and the 414 downregulated genes (blue).

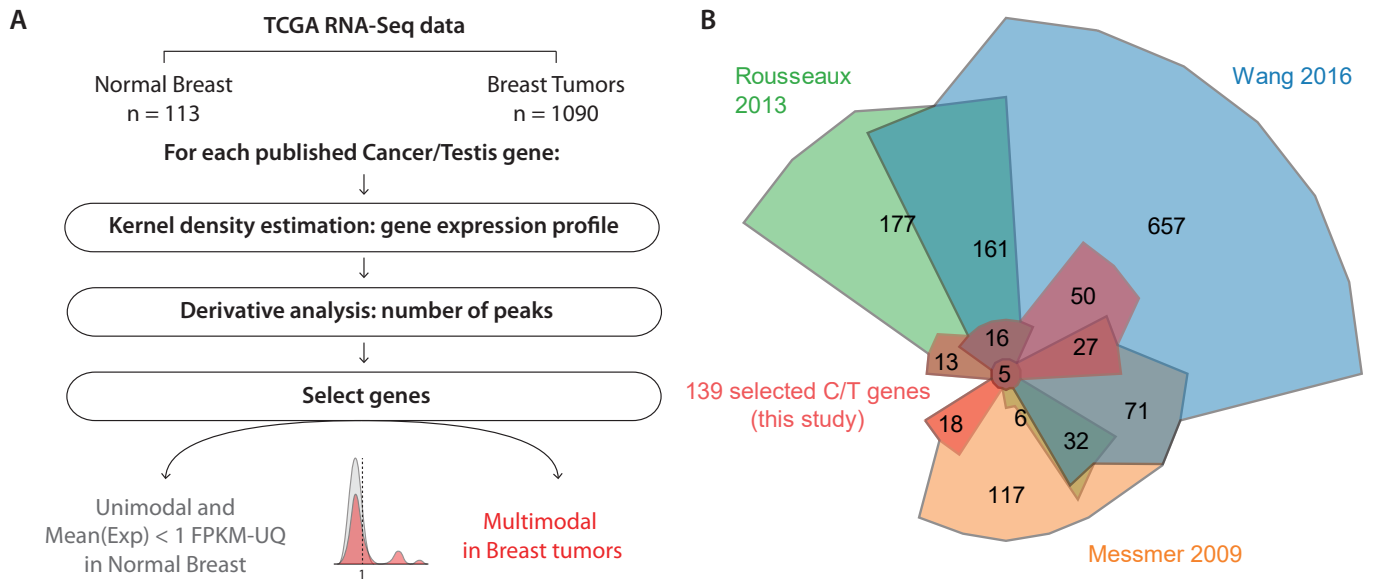


Figure 1: A custom bioinformatic screen identifies 139 Cancer/Testis genes abnormally expressed in breast tumors

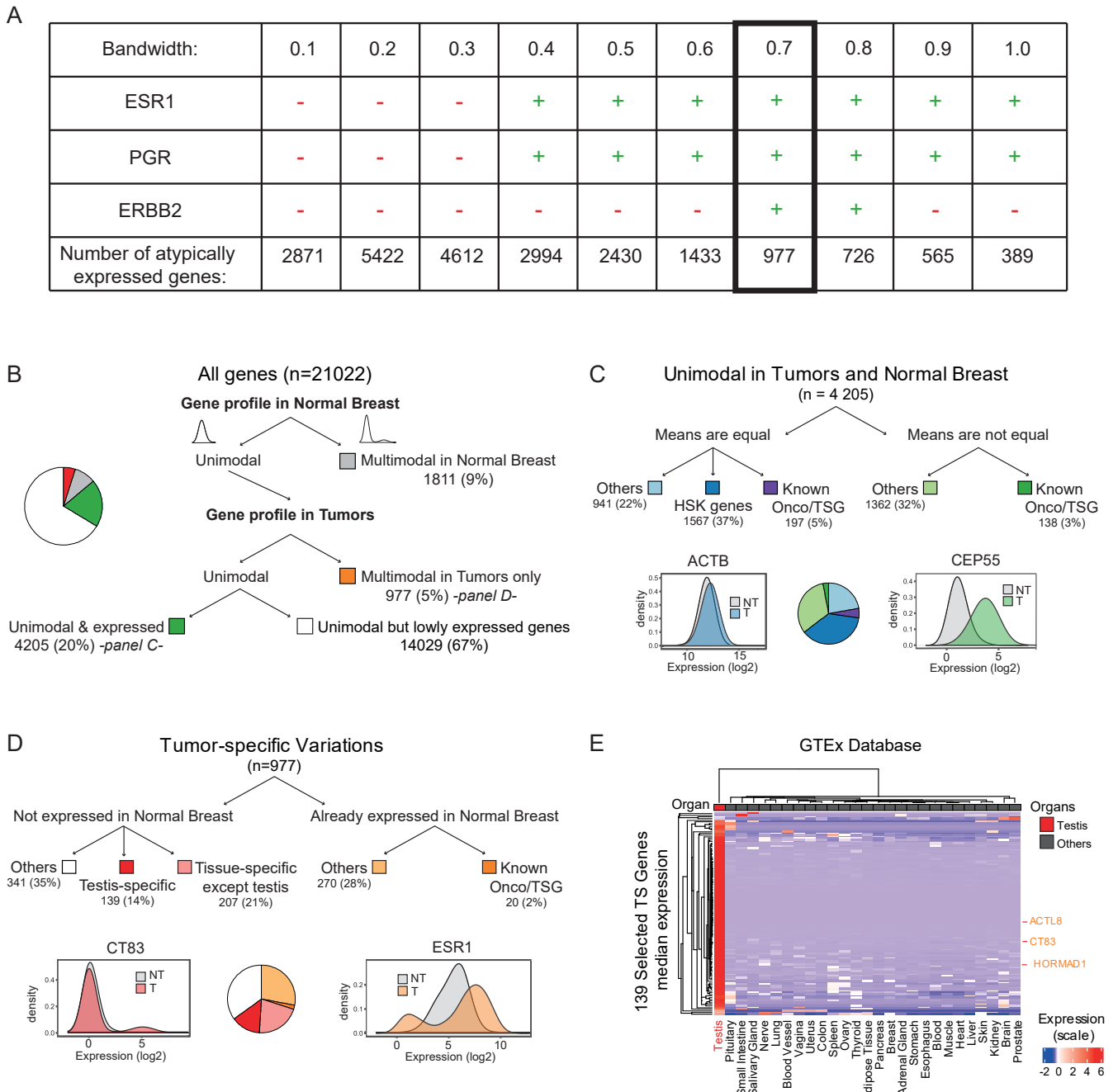


Figure S1: Optimizing parameters for the bioinformatic screen, further uses and validations

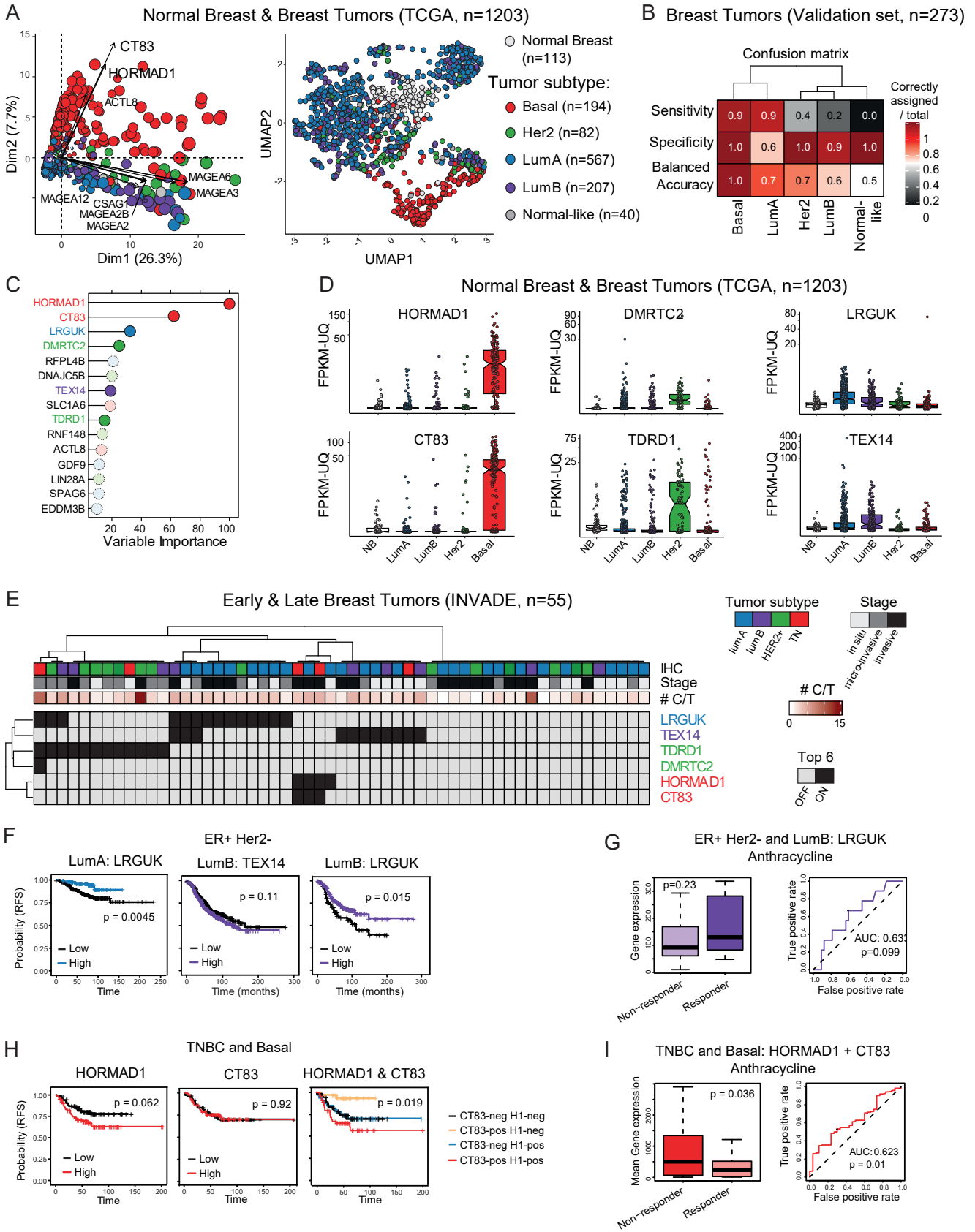


Figure 2: The activation of specific C/T genes is predictive of tumor subtype, occurs early during tumorigenesis, and is associated with prognosis

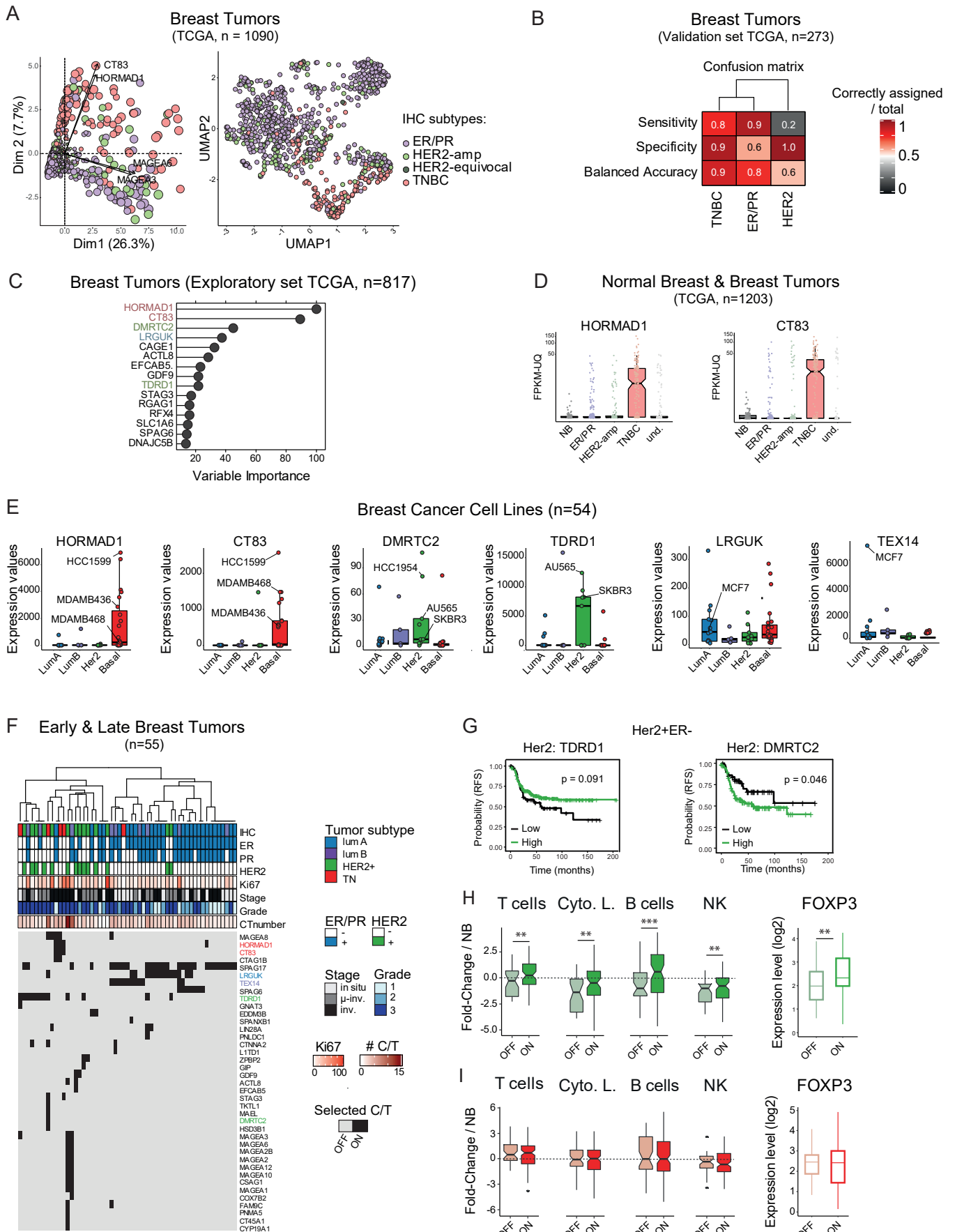


Figure S2 Examination of marker expression in tumors classified by IHC and in tumor cell lines. Expression in early tumors and association with survival.

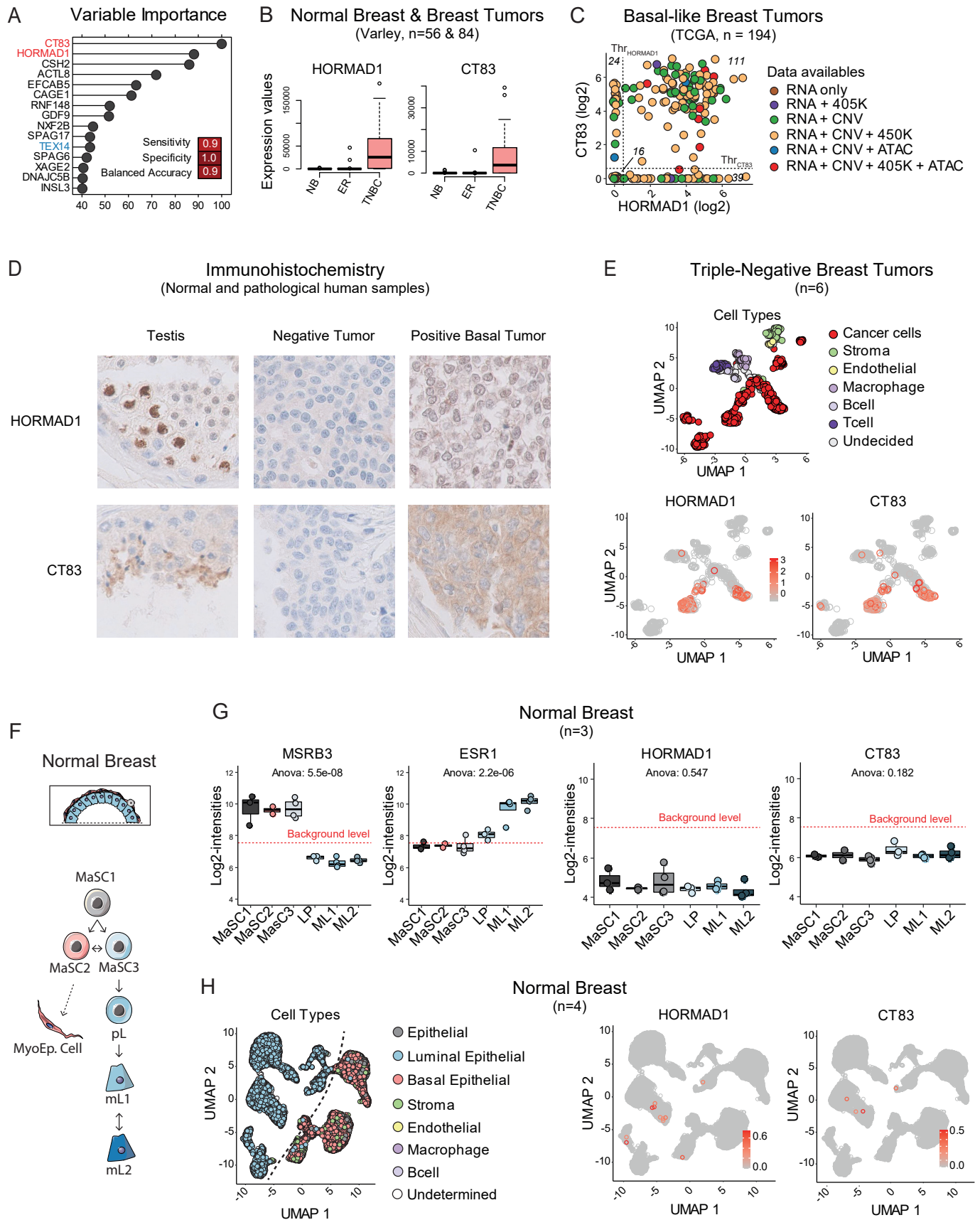


Figure 3: HORMAD1 and CT83 are expressed specifically by cancer cells, however scRNA-seq reveals rare HORMAD1+ / CT83+ luminal progenitor cells in healthy mammary gland

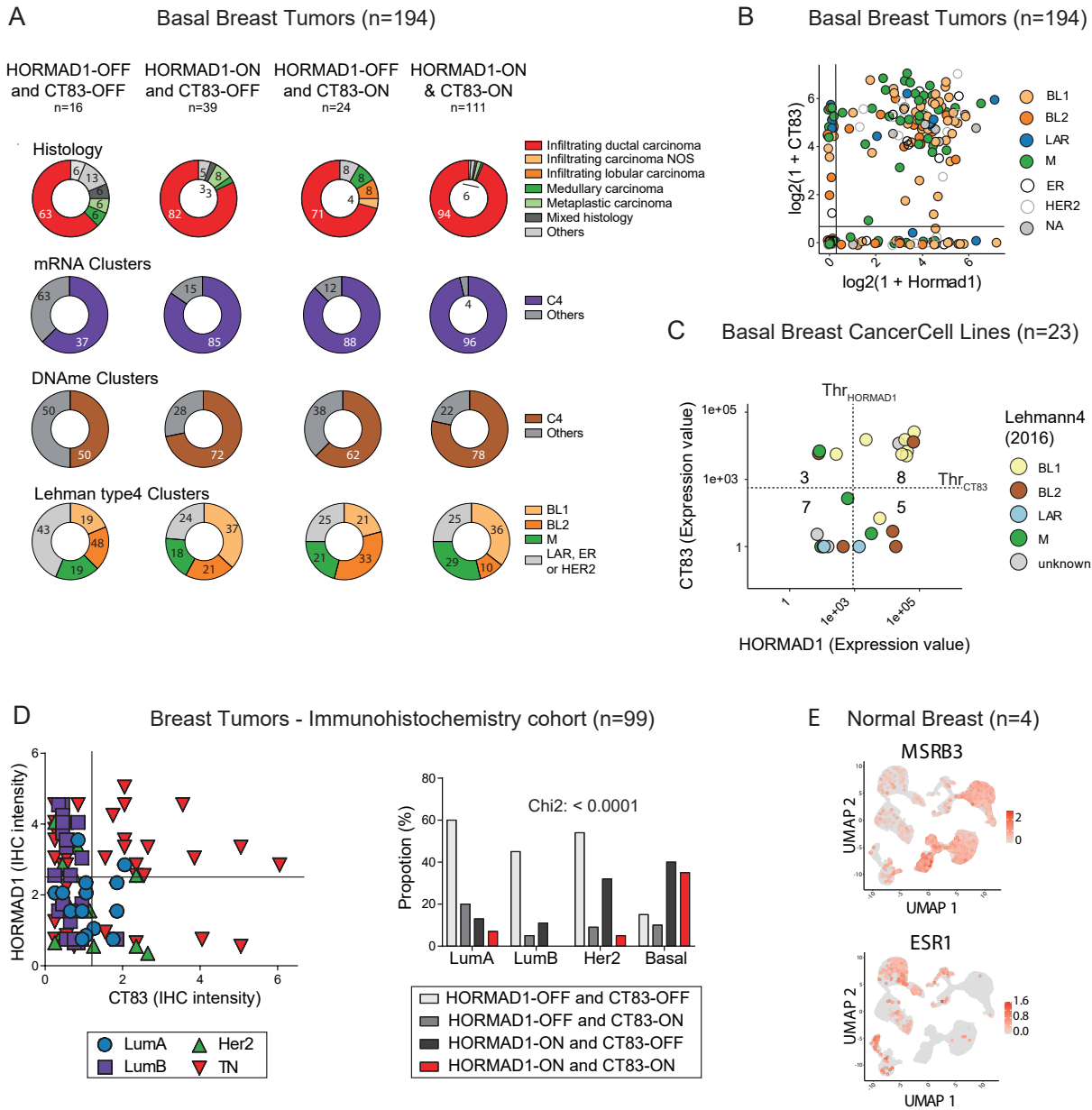


Figure S3: Characteristics of HORMAD1/CT83-positive basal tumors and cell lines, validation by IHC

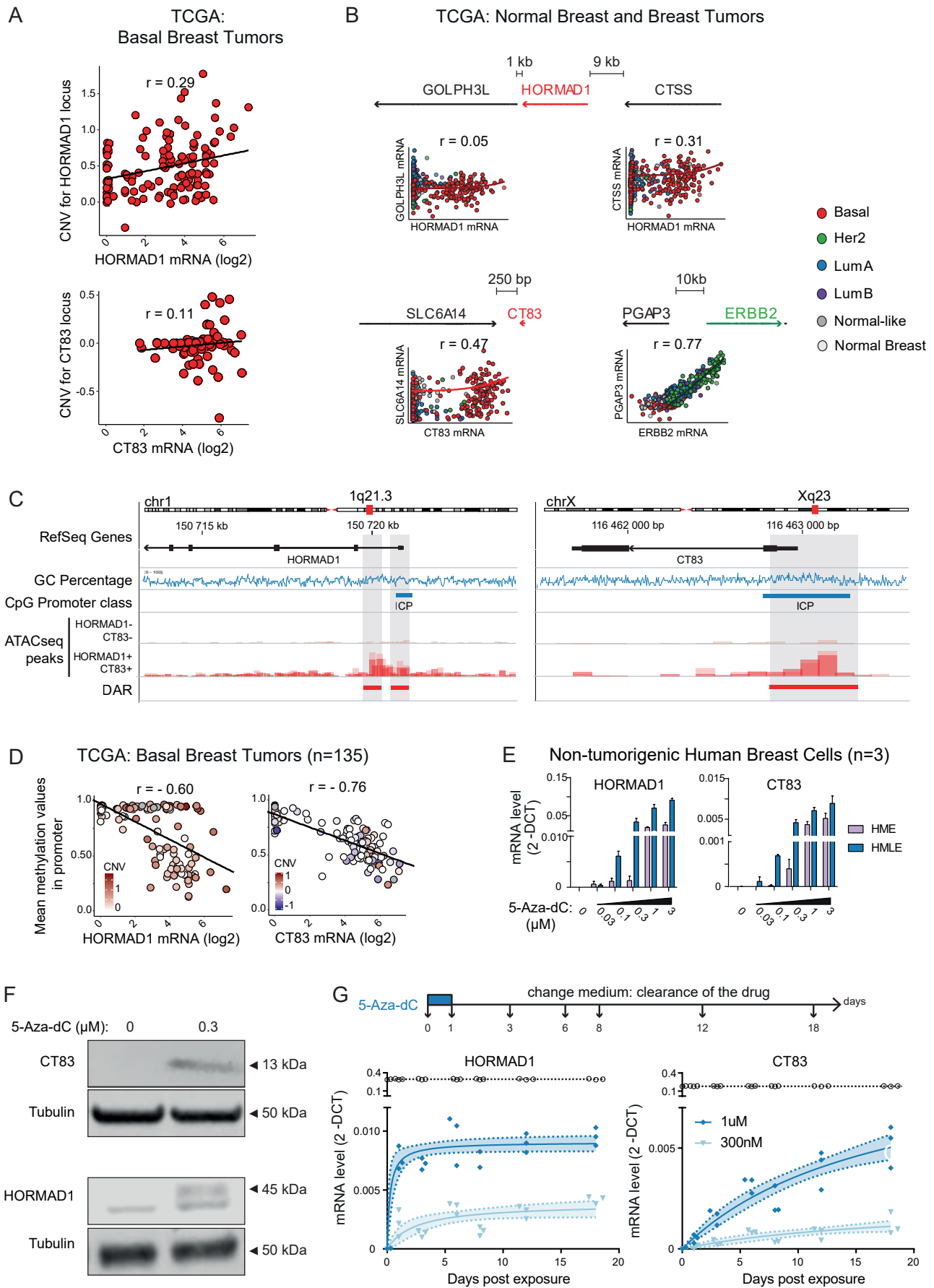


Figure 4: HORMAD1 & CT83 expressions are epigenetically regulated, with an essential contribution of DNA methylation

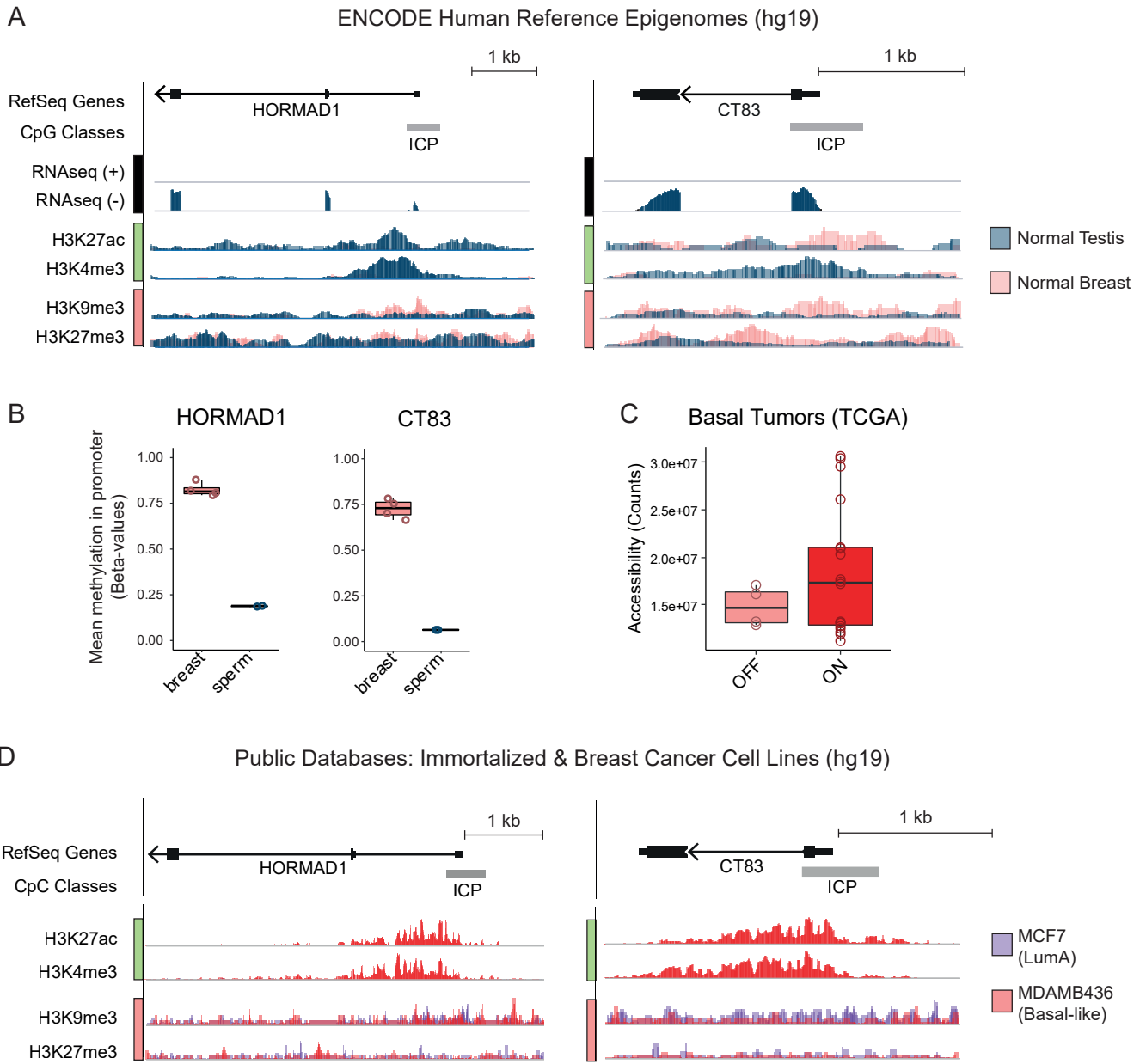


Figure S4: Epigenetic landscapes of the HORMAD1 and CT83 genes

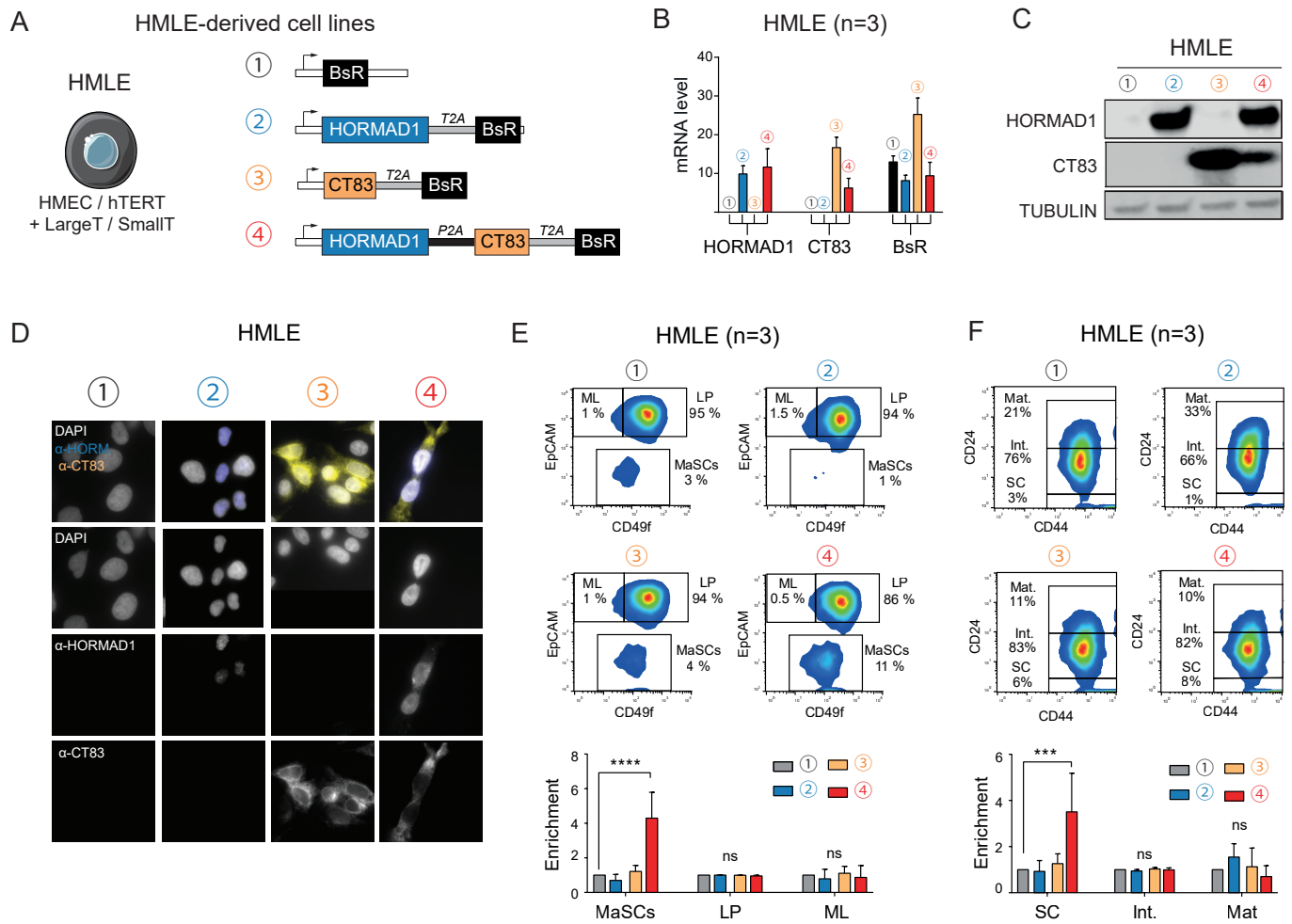


Figure 5: HORMAD1 & CT83 act synergistically to promote aggressive features in non-transformed mammary cells

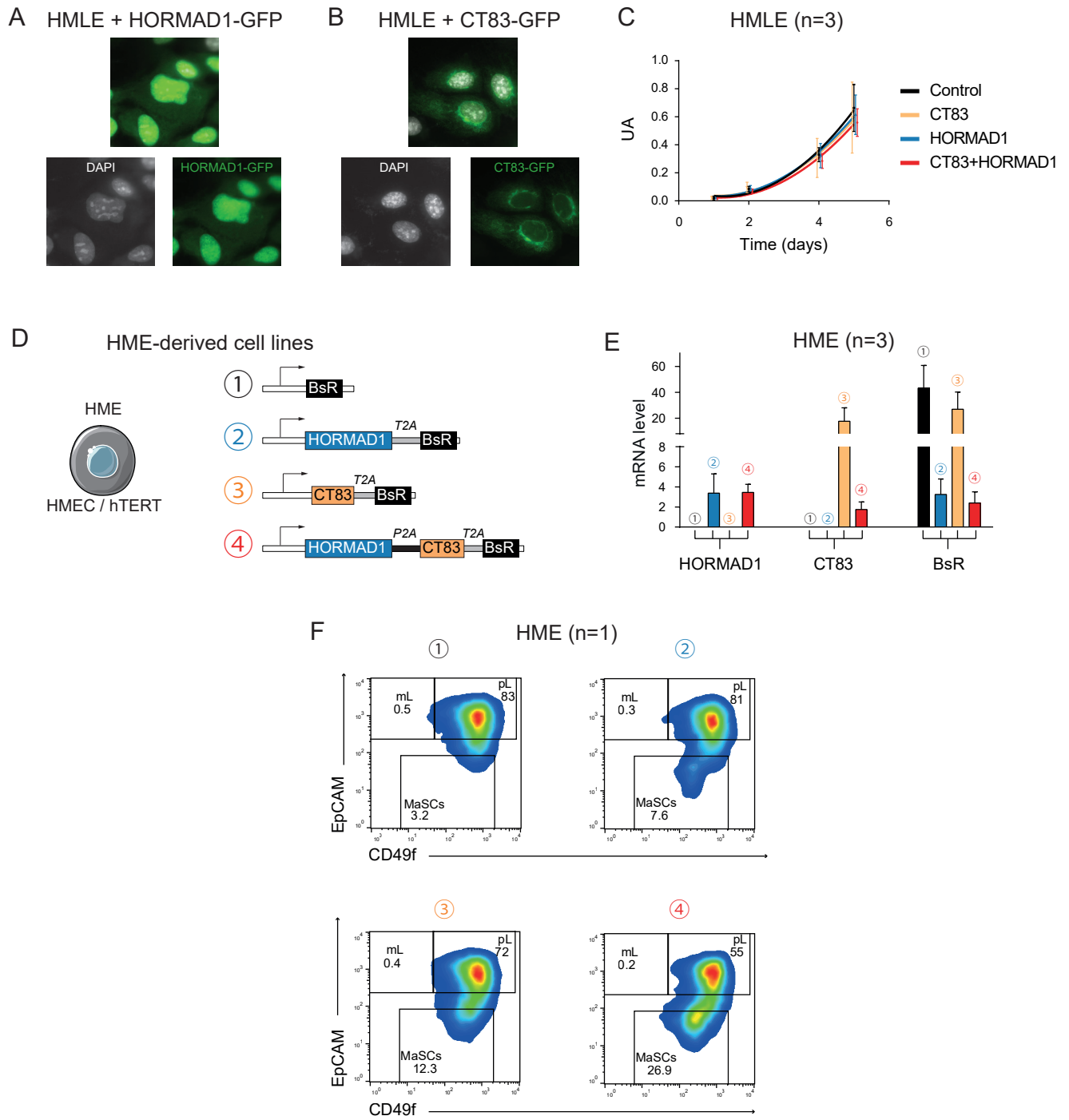
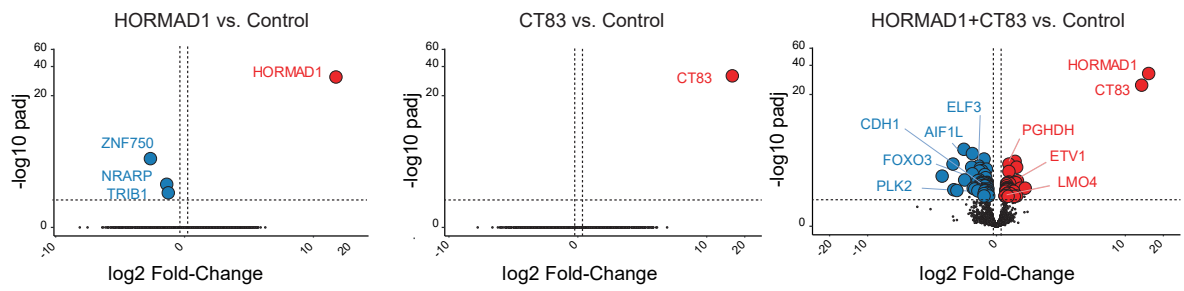


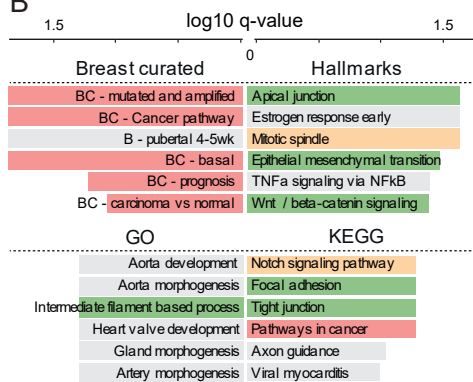
Figure S5: Additional experiments on HORMAD1 and CT83 in cultured breast cells

A

RNA-seq HMLE: Differential gene expression analysis



B



C

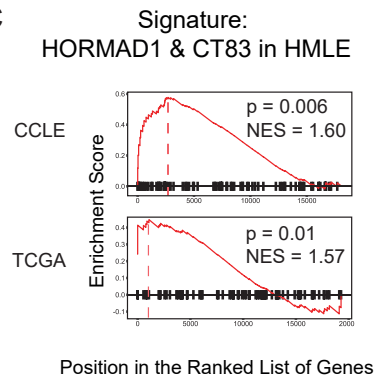
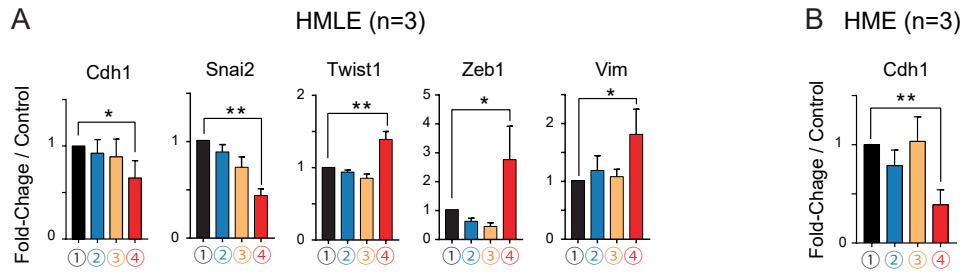
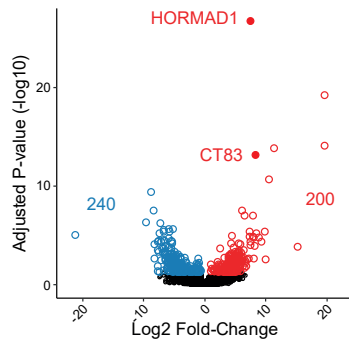


Figure 6: HORMAD1 & CT83-induced phenotype in HMLE cell line is also found in double-positive basal breast tumors



C CCLE RNA-seq: Basal Breast Cancer Cell Lines
HORMAD1 & CT83-positive vs. -negatives



D TCGA RNA-seq: Basal Breast Tumors
HORMAD1 & CT83-positive vs. -negatives

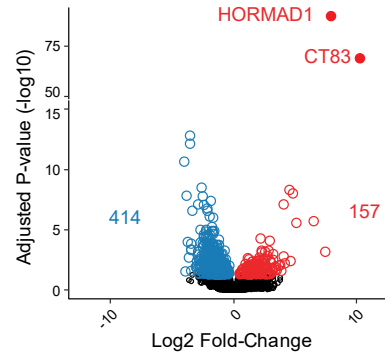


Figure S6: Validation of the "HORMAD1+CT83" signature, additional analyses in tumors and cell lines

TS1: 139 Selected Cancer/Testis genes

gene_id	gene_id	entrezgene	ensembl_
CRISP1 167	CRISP1	167	ENSG00000124812
AQP5 362	AQP5	362	ENSG00000161798
CSH2 1443	CSH2	1443	ENSG00000213218
CTAG1B 1485	CTAG1A1	485	ENSG00000183678
CTAG1B 1485.1	CTAG1B1	485	ENSG00000184033
CTNNA2 1496	CTNNA2	1496	ENSG00000066032
CYP19A1 1588	CYP19A1	1588	ENSG00000137869
DMP1 1758	DMP1	1758	ENSG00000152592
GDF9 2661	GDF9	2661	ENSG00000164404
GIP 2695	GIP	2695	ENSG00000159224
GPX5 2880	GPX5	2880	ENSG00000224586
HSD3B1 3283	HSD3B1	3283	ENSG00000203857
IGFBP1 3484	IGFBP1	3484	ENSG00000146678
INSL3 3640	INSL3	3640	ENSG00000248099
INSL4 3641	INSL4	3641	ENSG00000120211
KRT33B 3884	KRT33B	3884	ENSG00000131738
MAGEA1 4100	MAGEA1	4100	ENSG00000198681
MAGEA2 4101	MAGEA2	4101	ENSG00000184750
MAGEA2 4101.1	MAGEA2B	4101	ENSG00000183305
MAGEA3 4102	MAGEA3	4102	ENSG00000221867
MAGEA6 4105	MAGEA3	4105	ENSG00000221867
MAGEA6 4105.1	MAGEA6	4105	ENSG00000197172
MAGEA8 4107	MAGEA8	4107	ENSG00000156009
MAGEA10 4109	MAGEA10	4109	ENSG00000124260
MAGEA12 4111	MAGEA12	4111	ENSG00000213401
MAGEB2 4113	MAGEB2	4113	ENSG00000099399
NMBR 4829	NMBR	4829	ENSG00000135577
PENK 5179	PENK	5179	ENSG00000181195
PSG9 5678	PSG9	5678	ENSG00000183668
PSG11 5680	PSG11	5680	ENSG00000243130
PSG11 5680.1	PSG3	5680	ENSG00000221826
RFX4 5992	RFX4	5992	ENSG00000111783
SLC1A6 6511	SLC1A6	6511	ENSG00000105143
SSX1 6756	SSX1	6756	ENSG00000126752
SSX5 6758	SSX5	6758	ENSG00000165583

AURKC 6795	AURKC	6795	ENSG00000105146
TNP1 7141	TNP1	7141	ENSG00000118245
DNALI1 7802	DNALI1	7802	ENSG00000163879
TKTL1 8277	TKTL1	8277	ENSG00000007350
PAGE1 8712	PAGE1	8712	ENSG00000068985
XAGE2 9502	XAGE2B	9502	ENSG00000155622
XAGE2 9502.1	XAGE2	9502	ENSG00000185751
XAGE1D 9503	XAGE1A	9503	ENSG00000204379
XAGE1D 9503.1	XAGE1E	9503	ENSG00000204375
XAGE1D 9503.2	XAGE1D	9503	ENSG00000204376
XAGE1D 9503.3	XAGE1C	9503	ENSG00000183461
XAGE1D 9503.4	XAGE1B	9503	ENSG00000204382
PAGE4 9506	PAGE4	9506	ENSG00000101951
SPAG6 9576	SPAG6	9576	ENSG00000077327
SSX3 10214	SSX3	10214	ENSG00000165584
SSX3 10214.1	SSX5	10214	ENSG00000165583
STAG3 10734	STAG3	10734	ENSG00000066923
CAPN11 11131	CAPN11	11131	ENSG00000137225
SPO11 23626	SPO11	23626	ENSG00000054796
TMEFF2 23671	TMEFF2	23671	ENSG00000144339
AIPL1 23746	AIPL1	23746	ENSG00000129221
CABYR 26256	CABYR	26256	ENSG00000154040
ZBTB32 27033	ZBTB32	27033	ENSG00000011590
RBMXL2 27288	RBMXL2	27288	ENSG00000170748
VCX2 51480	VCX2	51480	ENSG00000177504
VCX3A 51481	VCX3A	51481	ENSG00000169059
L1TD1 54596	L1TD1	54596	ENSG00000240563
NXF2 56001	NXF2	56001	ENSG00000185554
NXF2 56001.1	NXF2B	56001	ENSG00000185945
TEX14 56155	TEX14	56155	ENSG00000121101
TEX11 56159	TEX11	56159	ENSG00000120498
TDRD1 56165	TDRD1	56165	ENSG00000095627
ANKRD7 56311	ANKRD7	56311	ENSG00000106013
TRIM49 57093	TRIM49	57093	ENSG00000168930
SPINLW1 57119	EPPIN	57119	ENSG00000101448
RGAG1 57529	RGAG1	57529	ENSG00000243978
DMRTC2 63946	DMRTC2	63946	ENSG00000142025
NEUROG2 63973	NEUROG2	63973	ENSG00000178403
EDDM3B 64184	EDDM3B	64184	ENSG00000181552

C19orf57 79173	C19orf57	79173	ENSG00000132016
BCL2L14 79370	BCL2L14	79370	ENSG00000121380
LIN28A 79727	LIN28A	79727	ENSG00000131914
LIN28A 79727.1	LIN28AP1	79727	ENSG00000213120
ACTL8 81569	ACTL8	81569	ENSG00000117148
TEX101 83639	TEX101	83639	ENSG00000131126
HORMAD1 84072	HORMAD1	84072	ENSG00000143452
DSCR8 84677	DSCR8	84677	ENSG00000198054
NAA11 84779	NAA11	84779	ENSG00000156269
MAEL 84944	MAEL	84944	ENSG00000143194
DNAJC5B 85479	DNAJC5B	85479	ENSG00000147570
FATE1 89885	FATE1	89885	ENSG00000147378
PAGE5 90737	PAGE5	90737	ENSG00000158639
TDRD12 91646	TDRD12	91646	ENSG00000173809
SYCE1 93426	SYCE1	93426	ENSG00000171772
CGB5 93659	CGB5	93659	ENSG00000189052
CGB5 93659.1	CGB	93659	ENSG00000104827
CGB5 93659.2	CGB8	93659	ENSG00000213030
PNMA5 114824	PNMA5	114824	ENSG00000198883
CATSPER1 117144	CATSPER1	117144	ENSG00000175294
ZPBP2 124626	ZPBP2	124626	ENSG00000186075
C17orf64 124773	C17orf64	124773	ENSG00000141371
ZDHHC19 131540	ZDHHC19	131540	ENSG00000163958
ZFP42 132625	ZFP42	132625	ENSG00000179059
NOBOX 135935	NOBOX	135935	ENSG00000106410
LRGUK 136332	LRGUK	136332	ENSG00000155530
DCAF12L1 139170	DCAF12L1	139170	ENSG00000198889
MAGEB16 139604	MAGEB16	139604	ENSG00000189023
RPL10L 140801	RPL10L	140801	ENSG00000165496
C20orf152 140894	CNBD2	140894	ENSG00000149646
C10orf82 143379	C10orf82	143379	ENSG00000165863
LYPD4 147719	LYPD4	147719	ENSG00000183103
FAM187B 148109	FAM187B	148109	ENSG00000177558
PNLDC1 154197	PNLDC1	154197	ENSG00000146453
CSAG1 158511	CSAG1	158511	ENSG00000198930
FMR1NB 158521	FMR1NB	158521	ENSG00000176988
FSIP1 161835	FSIP1	161835	ENSG00000150667
ADAD2 161931	ADAD2	161931	ENSG00000140955
RNF133 168433	RNF133	168433	ENSG00000188050

XAGE3 170626	XAGE3	170626	ENSG00000171402
XAGE5 170627 X	AGE5	170627	ENSG00000171405
COX7B2 170712	COX7B2	170712	ENSG00000170516
FAM9C 171484	FAM9C	171484	ENSG00000187268
SPAG17 200162	SPAG17	200162	ENSG00000155761
CXorf61 203413	CT83	203413	ENSG00000204019
PAGE2 203569	PAGE2	203569	ENSG00000234068
C18orf20 221241	LINC00305	221241	ENSG00000179676
C16orf73 254528	MEIOB	254528	ENSG00000162039
WFDC11 259239	WFDC11	259239	ENSG00000180083
ODF3L2 284451	ODF3L2	284451	ENSG00000181781
CAGE1 285782	CAGE1	285782	ENSG00000164304
TMEM95 339168	TMEM95	339168	ENSG00000182896
COX8C 341947	COX8C	341947	ENSG00000187581
GNAT3 346562	GNAT3	346562	ENSG00000214415
CXorf66 347487	CXorf66	347487	ENSG00000203933
C12orf42 374470	C12orf42	374470	ENSG00000179088
EFCAB5 374786	EFCAB5	374786	ENSG00000176927
RNF148 378925	RNF148	378925	ENSG00000235631
TSPYL6 388951	TSPYL6	388951	ENSG00000178021
C2orf78 388960	C2orf78	388960	ENSG00000187833
PAGE2B 389860	PAGE2B	389860	ENSG00000238269
BCAR4 400500	BCAR4	400500	ENSG00000262117
C1orf141 400757	C1orf141	400757	ENSG00000203963
C4orf40 401137	C4orf40	401137	ENSG00000187533
VCX3B 425054	VCX3B	425054	ENSG00000205642
LRRC52 440699	LRRC52	440699	ENSG00000162763
CT45A4 441520	CT45A4	441520	ENSG00000228836
CT45A4 441520.1	CT45A6	441520	ENSG00000226907
CT45A4 441520.2	CT45A2	441520	ENSG00000242185
RFPL4B 442247	RFPL4B	442247	ENSG00000251258
SPANXN5 494197	SPANXN5	494197	ENSG00000204363
CT45A1 541466	CT45A3	541466	ENSG00000232417
CT45A1 541466.1	CT45A1	541466	ENSG00000232478
RAD21L1 642636	RAD21L1	642636	ENSG00000244588
RHOXF2B 727940	RHOXF2B	727940	ENSG00000203989
CT45A2 728911	CT45A4	728911	ENSG00000228836
CT45A2 728911.1	CT45A2	728911	ENSG00000242185
CT45A2 728911.2	CT45A1	728911	ENSG00000232478

CT45A2 728911.3	CT45A3	728911	ENSG00000232417
GAGE8 100101629	GAGE2E	100101629	ENSG00000205775
GAGE8 100101629.1	GAGE2D	100101629	ENSG00000240257
SPANXB2 100133171	SPANXB1	100133171	ENSG00000235604
SPANXB2 100133171.1	SPANXB2	100133171	ENSG00000227234

TS2: Basal-like tumor characteristics

		Both_OFF		HORMAD1_Only		CT83_Only		Both_ON		Chi2	AOV
N		16		39		24		111		190	
ER	Negative	14	87,50	35	89,74	22	91,67	96	86,49		
	Positive	2	12,50	2	5,13	2	8,33	9	8,11		
	NA	0	0,00	2	5,13	0	0,00	6	5,41	ns	
PR	Negative	12	75,00	34	87,18	24	100,00	99	89,19		
	Positive	4	25,00	4	10,26	0	0,00	7	6,31		
	NA	0	0,00	1	2,56	0	0,00	5	4,50	ns	
HER2	Negative	16	100,00	37	94,87	23	95,83	102	91,89		
	Positive	0	0,00	1	2,56	1	4,17	4	3,60		
	NA	0	0,00	1	2,56	0	0,00	5	4,50	ns	
Stage	Stage I-II	10	62,50	32	82,05	19	79,17	95	85,59		
	Stage III-IV	6	37,50	7	17,95	4	16,67	13	11,71		
	NA	0	0,00	0	0,00	1	4,17	3	2,70	ns	
Histological Type	Infiltrating ductal carcinoma	10	62,50	32	82,05	17	70,83	104	93,69		
	Infiltrating carcinoma NOS	0	0,00	0	0,00	1	4,17	0	0,00		
	Infiltrating lobular carcinoma	0	0,00	0	0,00	2	8,33	0	0,00		
	Medullary carcinoma	1	6,25	1	2,56	2	8,33	1	0,90		
	Metaplastic carcinoma	1	6,25	3	7,69	0	0,00	2	1,80		
	Mixed histology	1	6,25	1	2,56	0	0,00	1	0,90		
	Other	2	12,50	2	5,13	2	8,33	2	1,80		
	NA	1	6,25	0	0,00	0	0,00	1	0,90	NA	
Histological Type 2	Infiltrating ductal carcinoma	10	62,50	32	82,05	17	70,83	104	93,69		
	Others	5	31,25	7	17,95	7	29,17	6	5,41		
	NA	1	6,25	0	0,00	0	0,00	1	0,90	0.0006852	
Age at diagnosis	mean +-sd	59.2 ± 10.8		54.6 ± 11.4		56.9 ± 12.1		55.7 ± 12.7			ns
Initial weight		280 ± 248		326 ± 259		398 ± 309		289 ± 225			ns
Number of positive lymph nodes by he	mean +-sd	1.1 ± 1.8		1.7 ± 3.6		1.3 ± 2.3		1.3 ± 2.9			ns
Number of positive lymph nodes by he	NO	9	56,25	25	64,10	13	54,17	65	58,56		
	N1	3	18,75	6	15,38	6	25,00	28	25,23		
	N>1	2	12,50	6	15,38	3	12,50	10	9,01	ns	
Metastasis at diagnosis	NO	7	43,75	12	30,77	8	33,33	30	27,03		
	YES	0	0,00	1	2,56	0	0,00	2	1,80		
	NA	9	56,25	26	66,67	16	66,67	79	71,17	ns	
Lehman subtype	BL1	1	6,25	9	23,08	3	12,50	21	18,92		
	BL2	2	12,50	3	7,69	3	12,50	7	6,31		
	ER	2	12,50	2	5,13	2	8,33	8	7,21		
	HER2	3	18,75	5	12,82	1	4,17	13	11,71		
	IM	2	12,50	7	17,95	2	8,33	17	15,32		
	LAR	1	6,25	1	2,56	0	0,00	1	0,90		
	M	3	18,75	6	15,38	5	20,83	27	24,32		
	MSL	1	6,25	2	5,13	4	16,67	8	7,21		
NA	1	6,25	4	10,26	4	16,67	9	8,11	ns		
Lehman IV subtype	BL1	3	18,75	14	35,90	5	20,83	37	33,33		
	BL2	3	18,75	8	20,51	8	33,33	11	9,91		
	ER	2	12,50	2	5,13	2	8,33	8	7,21		
	HER2	3	18,75	5	12,82	1	4,17	13	11,71		
	LAR	2	12,50	2	5,13	3	12,50	5	4,50		
	M	3	18,75	7	17,95	5	20,83	30	27,03		
NA	0	0,00	1	2,56	0	0,00	7	6,31	ns		
Other	BL1	3	18,75	14		5		37			
	BL2	3	18,75	8		8		11			
	M	3		7		5		30			
	Other	7		9		6		26			
	NA	0	0,00	2	5,13	1	4,17	0	0,00	NA	
DNAmethylation Cluster	C1	1	6,25	0	0,00	0	0,00	1	0,90		
	C2	5	31,25	9	23,08	6	25,00	11	9,91		
	C3	1	6,25	0	0,00	1	4,17	12	10,81		
	C4	8	50,00	28	71,79	15	62,50	87	78,38		
	C5	1	6,25	0	0,00	1	4,17	0	0,00		
	C6	0	0,00	0	0,00	0	0,00	0	0,00		
	NA	0	0,00	2	5,13	1	4,17	0	0,00	NA	
DNA met 2	C4	8	50,00	28	71,79	15	62,50	87	78,38		
	Other	8	50,00	11	28,21	9	37,50	24	21,62	ns	
	C2	2	12,50	2	5,13	2	8,33	4	3,60		

mRNA Cluster	C3	1	6,25	1	2,56	1	4,17	0	0,00	NA
	C4	10	62,50	33	84,62	21	87,50	107	96,40	
	C7	3	18,75	3	7,69	0	0,00	0	0,00	
	NA	0	0,00	0	0,00	0	0,00	0	0,00	
mRNA 2	C4	10	62,50	33	84,62	21	87,50	107	96,40	0.0001769
	Other	6	37,50	6	15,38	3	12,50	4	3,60	
lncRNA Cluster	C1	1	6,25	0	0,00	2	8,33	1	0,90	na
	C2	3	18,75	11	28,21	5	20,83	36	32,43	
	C3	2	12,50	7	17,95	5	20,83	17	15,32	
	C4	1	6,25	2	5,13	0	0,00	2	1,80	
	C5	0	0,00	0	0,00	1	4,17	3	2,70	
	C6	1	6,25	3	7,69	5	20,83	11	9,91	
	NA	8	50,00	16	41,03	6	25,00	41	36,94	
miRNA Cluster	C1	0	0,00	0	0,00	0	0,00	1	0,90	NA
	C2	0	0,00	1	2,56	1	4,17	2	1,80	
	C3	2	12,50	2	5,13	2	8,33	3	2,70	
	C4	0	0,00	0	0,00	0	0,00	1	0,90	
	C5	4	25,00	4	10,26	0	0,00	2	1,80	
	C6	0	0,00	0	0,00	0	0,00	2	1,80	
	C7	10	62,50	29	74,36	21	87,50	95	85,59	
	NA	0	0,00	3	7,69	0	0,00	5	4,50	
CNV cluster	C1	0	0,00	0	0,00	0	0,00	1	0,90	ns
	C3	1	6,25	0	0,00	0	0,00	0	0,00	
	C4	9	56,25	27	69,23	17	70,83	78	70,27	
	C5	1	6,25	2	5,13	1	4,17	2	1,80	
	C6	5	31,25	8	20,51	5	20,83	16	14,41	
	NA	0	0,00	2	5,13	1	4,17	14	12,61	

Primer name	Sequence
PGK1-F	AGGATAAAGTCAGCCATGTGAG
PGK1-R	CACAGGAACTAAAAGGCAGGA
TBP-F	TGGCCCATAGTGATCTTTGC
TBP-R	TCCTAGAGCATCTCCAGCACA
HORMAD1-F	CAGTTGCAGAGGACTCCCAT
HORMAD1-R	CCATAAGCGCATTCTGGGAA
CT83-F	CGCCGCTTTCAGAGAAACAC
CT83-R	CCCGAGAGAGGTCGTAGACT
BLASTICIDIN-F	GACCTTGTGCAGAACTCGTG
BLASTICIDIN-F	AGGGCAGCAATTCACGAATC
TWIST1-F	GGC TCA GCT ACG CCT TCT C
TWIST1-R	CCT TCT CTG GAA ACA ATG ACA TCT
ZEB1-F	AGG GCA CAC CAG AAG CCA G
ZEB1-R	GAG GTA AAG CGT TTA TAG CCT CTA TCA
VIM-F	ATCCAAGTTTGCTGACCTCTGAG
VIM-R	AGGGACTGCACCTGTCTCCGGT
SNAI2-F	TGG TTG CTT CAA GGA CAC AT
SNAI2-R	GTT GCA GTG AGG GCA AGA A
CDH1-F	GGA ACTATGAAAAGTGGGCTTG
CDH1-R	AAATTGCCAGGCTCAATGAC
CDH2-F	CTTGTCAGGATCAGGTCT
CDH2-R	GAAGATACCAGTTGGAGGCT

Database Name	Author	Year	ID	URL
TCGA-BRCA	TCGA Research Network	2019		https://www.cancer.gov/tcga
TCGA-BRCA	TCGA Research Network	2019		https://www.cancer.gov/tcga
Ctdatabase	Gonzaga Almeida L.	2009	PMC2686577	http://www.cta.incc.br/
Testis-Specific / Placenta-Specific	Rousseaux S.	2013	PMC4818008	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/pmc/articles/PMC4818008/
C/T gene	Wang C	2016	PMC4737856	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/pmc/articles/PMC4737856/
Housekeeping genes	Eisenberg E & Levanon EY	2013	PMID: 23810203	https://www.tau.ac.il/~elieis/HKGL/
Oncogenes	UniProtKB	2021	https://www.uniprot.org/	https://www.uniprot.org/uniprot/ 'Proto-oncogene'
Tissue-specific genes	Kim P	2018	PMC5753286	http://zhaobioinfo.org/TissGDB
GTEx project	Carithers LJ	2015	PMC4675181	https://gtexportal.org/home/
GEO-BRCA	Varley KE	2014	GSE58135	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE58135
GEO-BRCA	Varley KE	2014	GSE58135	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE58135
CCL	Barretina J	2012	PMC3320027	https://portals.broadinstitute.org/ccle
INVADE	Vincent-Salomon A	2021	Unpublished yet	
EBI	Morel AP, Ginestier C	2017	E-MTAB-4145	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4145/files/
GEO-scBRCA	Cristea S	2018	GSE118389	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE118389
GEO-scNormal Mammary gland	Kessenbrock K	2018	GSE113197	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE113197
Human genome hg19				
TCGA-BRCA	TCGA Research Network	2019		https://www.cancer.gov/tcga
TCGA-BRCA	TCGA Research Network	2019		https://www.cancer.gov/tcga
ENCODE Testis Breast	The ENCODE Project Consortium	2012	PMC3439153	https://www.encodeproject.org/
GEO MCF10A MCF7 MDA-MB-436	Xi Y, Shi X, Li W, Allton K	2017	GSE85158	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE85158
GEO MDA-MB-436	Hatice O, Christina LS	2019	GSE114964	https://www.ncbi.nlm.nih.gov.insb.bib.cnrs.fr/geo/query/acc.cgi?acc=GSE114964

Sample Type	Sample size	Data Type	Data format	Plateform
Breast Tumors		1109 RNAseq	counts	Illumina HiSeq
Normal mammary gland		113 RNAseq	counts	Illumina HiSeq
Gene List	276 genes			
Gene List	411 genes			
Gene List	1019 genes			
Gene List	3804 genes			
Gene List	560 genes			
Gene List	631 genes			
Normal Tissues		1927 RNAseq	Processed TPM	Illumina HiSeq
Breast Tumors		84 RNAseq	Processed counts	Illumina HiSeq
Normal mammary gland		56 RNAseq	Processed counts	Illumina HiSeq
Breast Cancer Cell lines		69 RNAseq	Processed counts	Illumina HiSeq
Breast Tumors		55 RNAseq	Processed counts	Illumina HiSeq
Normal mammary gland	9 samples, pooled in 3 replicats	Expression microarray	raw data CEL	Affymetrix GeneChip Human Gene 1.0 ST Array
Breast Tumors	6 individus - 1 534 cells	scrRNAseq	raw data FASTQ	Illumina Genome Analyze
Normal mammary gland	4 individus - 24 646 cells	scrRNAseq	Processed FPKM	Illumina HiSeq 2500/4000
Normal			Processed beta values	
Breast Tumors		135 DNA methylation	Processed	HumanMethylation450 BeadChip
Breast Tumors		CNV data	Processed	Affymetrix SNP 6.0 array
Normal Tissues		4 ChIP-seq	Processed log2 FoldChange BigWig	ENCODE Processing Pipeline
Normal mammary and breast cancer cell line		6 ChIP-seq	Processed - bigWig	Illumina HiSeq 2000
Breast Cancer Cell line		2 ATACseq	Processed - bigWig	Illumina HiSeq 2500

2. Résultats supplémentaires

En parallèle de la rédaction de cet article, j'ai pu effectuer quelques expériences supplémentaires dont les résultats nous permettront de spéculer sur les mécanismes permettant la coopération d'HORMAD1 et de CT83 et leur activation.

Un résultat non commenté dans l'article est la déplétion massive des cellules souches mammaires dans les HMLE, après induction d'HORMAD1. Cet effet est d'intensité variable selon les expériences, et semblent s'affirmer à mesure que l'expérience de cytométrie est effectué à distance de l'infection et la sélection des cellules (1-2 semaines). Les 4 lignées ont été traitées en parallèle, elles ne présentent pas de différence de prolifération, donc le nombre de cycles cellulaires effectué par les 4 lignées durant ce temps est comparable. J'ai voulu utiliser les constructions GFP (**FIGURE 4 A-D**), afin de mesurer si cette interaction entre HORMAD1 seul et le compartiment cellules souches est reproductible dans un système vectoriel indépendant. Les résultats sont présentés dans la figure 1 : bien que la proportion initiale de cellules souches soit très faible dans ces expériences, cette population disparaît complètement lorsque l'on exprime HORMAD1 dans les HMLE (**FIGURE 4E**) comme dans les HME (**FIGURE 5C**). Des expériences supplémentaires sont nécessaires pour explorer une éventuelle toxicité sélective de l'expression d'HORMAD1 dans les cellules souches mammaires ; toxicité qui serait éventuellement contre-carrée par la co-expression de CT83. A ce stade, ces expériences sont encore très préliminaires, mais nous indiquent un mécanisme éventuel venant soutenir cette association préférentielle entre HORMAD1 et CT83 dans les tumeurs du sein basal-like.

Nous avons également commencé à développer des approches fonctionnelles permettant de valider l'implication de la co-expression d'HORMAD1 et de CT83 dans l'EMT (**FIGURE 4F**). Ces approches semblent valider l'effet de la co-expression d'HORMAD1 et de CT83 dans la migration, mais demandent à être reproduites. Je souhaiterais également intégrer les lignées HORMAD1 seul, et CT83 seul, à ces analyses.

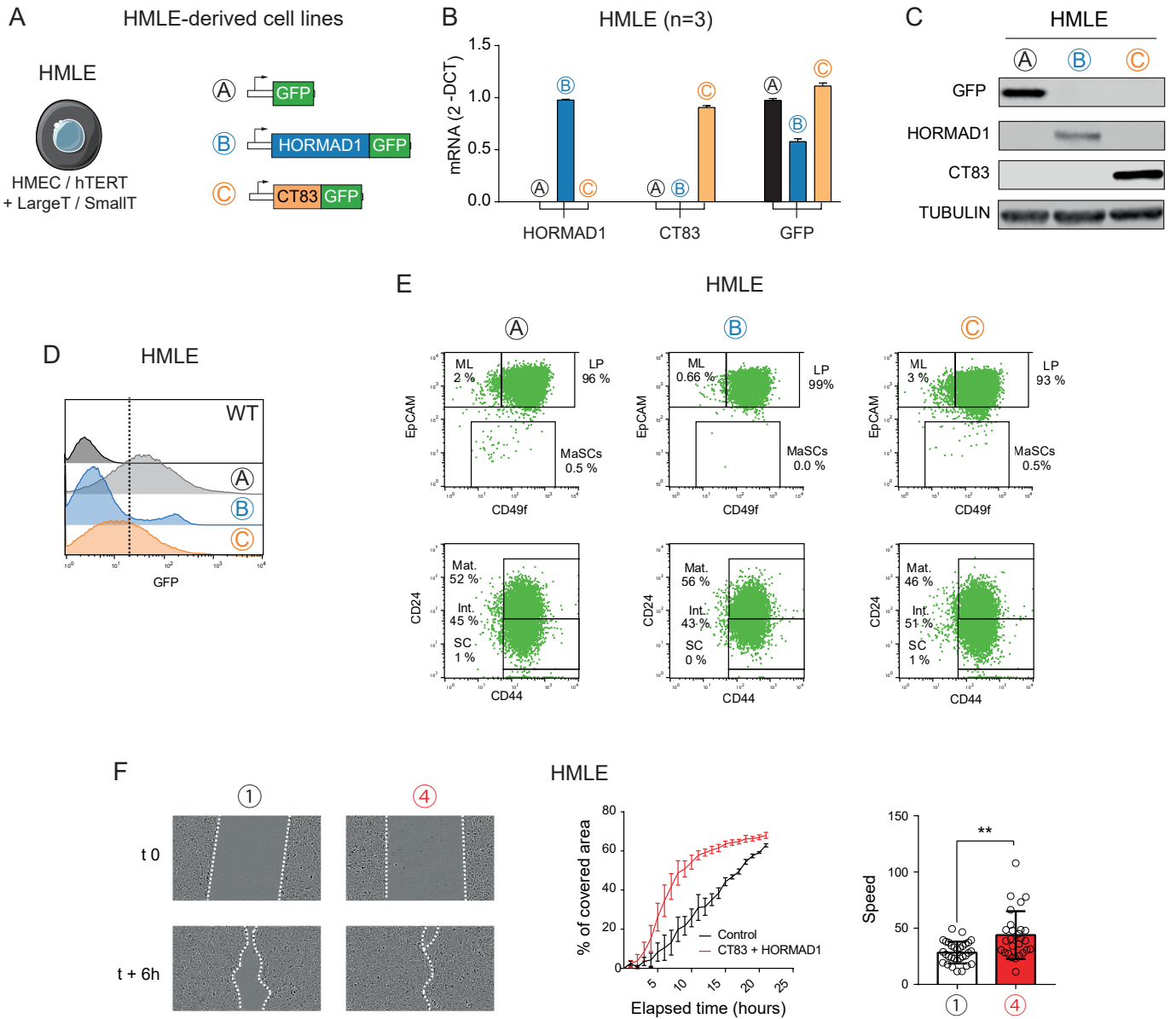


Figure 53 : Premières analyses fonctionnelles du rôle de HORMAD1 et CT83 dans la lignée HMLE

A. Modèle cellulaire et vecteurs lentiviraux utilisés pour exprimer HORMAD1 et / ou CT83. La GFP est fusionnée en N-terminale des protéines.

B. Expression d'HORMAD1, de CT83 et du gène rapporteur codant pour la GFP, évaluées par RTqPCR, dans les lignées dérivées de HMLE. Les données sont représentées en tant que moyenne +/- sd (n = 3 expériences indépendantes)

C. Expression de la GFP, d'HORMAD1 et de CT83 évaluée par Western Blot dans les lignées dérivées de HMLE.

D. Expression de la GFP, évaluées par FACS après infection des HMLE.

E. Haut: Analyse par cytométrie des marqueurs de surface EpCAM et CD49f dans les lignées dérivées de HMLE. Cellules luminales matures (ML, EpCAM+CD49f-), Progéniteurs luminaux (PL, EpCAM+CD49f+), Cellules souches mammaires (MaSC, EpCAM-CD49flow). Bas: Analyse par cytométrie des marqueurs de surface CD24 et CD44 dans les lignées dérivées de HMLE. Cellules matures (Mat, CD44+CD24high), intermédiaires (Int, CD44+CD24low), Cellules souches (SC, CD44+CD24low).

F. De gauche à droite: une expérience représentative de wound-healing en contraste de phase (4X) ; évolution de la blessure au cours du temps (moyenne +/-sd, n= 3 expériences indépendantes) ; vitesse moyenne de migration.

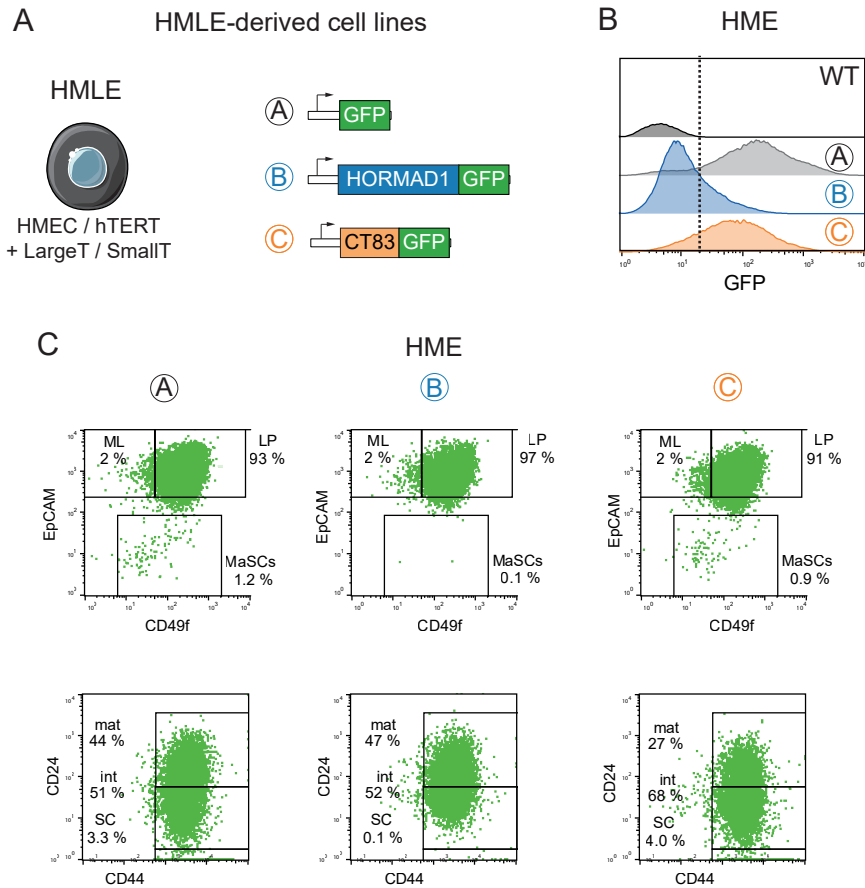


Figure 54 : Premières analyses fonctionnelles du rôle de HORMAD1 et CT83 dans la lignée HME

A. Modèle cellulaire et vecteurs lentiviraux utilisés pour exprimer HORMAD1 et / ou CT83. La GFP est fusionnée en N-terminale des protéines.

B. Expression de la GFP, évaluées par FACS après infection des HME.

C. Haut: Analyse par cytométrie des marqueurs de surface EpcAM et CD49f dans les lignées dérivées de HME. Cellules luminales matures (ML, EpcAM+CD49f-), Progéniteurs luminaux (LP, EpcAM+CD49f+), Cellules souches mammaires (MaSC, EpcAM-CD49flow). Bas: Analyse par cytométrie des marqueurs de surface CD24 et CD44 dans les lignées dérivées de HME. Cellules matures (Mat, CD44+CD24high), intermédiaires (Int, CD44+CD24low), Cellules souches (SC, CD44+CD24low).

Nous souhaitons identifier les mécanismes qui influencent la déméthylation d'HORMAD1 et de CT83 dans les cellules cancéreuses ou leurs progéniteurs. Pour cela, nous avons tiré profit d'une étude exploratoire réalisée en collaboration avec Gaël Christofari et Claude Phillippe (IRCAN) : elle repose sur l'extraction de la base de données GEO de toutes les analyses réalisées par puces 450K sur du matériel humain, que ce soit des lignées cellulaires, des organoïdes ou des échantillons de tissu. Ces données ont été homogénéisées et normalisées, les metadonnées associées ont été extraites et nettoyées afin d'identifier les tissus, types cellulaires, conditions expérimentales, maladies, etc... correspondant à chaque échantillon. Cette vaste base de données permet d'évaluer de façon non biaisée les facteurs corrélant avec l'hypométhylation des CpG impliqués dans la régulation de l'expression d'HORMAD1 et de CT83, tels que nous les avons identifiés dans les cancers du sein.

Gaël Christofari et Claude Phillippe ont extrait pour nous les données relatives aux sondes détectant des CpG dans les régions promotrices d'HORMAD1 et de CT83 : il s'agit de 7 sondes pour HORMAD1 et 8 pour CT83. Après avoir filtré les résultats, on obtient le niveau de méthylation de ces 15 CpG dans 13866 conditions. Ces données ont été projetées par UMAP (**FIGURE 6A**). On distingue un sous-groupe d'échantillons appartenant soit au système reproductif, soit issus de cancers du sein triple-négatif, soit ayant subi un traitement déméthylant. Ce résultat confirme la spécificité du statut de méthylation d'HORMAD1 et de CT83

On observe de plus une ségrégation claire des échantillons en deux clusters (**FIGURE 6A**). Après avoir nettoyé la base de donnée afin d'ajouter des informations relatives au sexe des donneurs, on s'aperçoit que cette ségrégation correspond à cette variable. En effet, le niveau de méthylation de certains CpG présents dans le promoteur de CT83 suit une distribution bimodale (**FIGURE 6B-C**), qui s'explique notamment par le sexe d'origine des échantillons : les femmes montrent fréquemment une hypométhylation de ces CpG. On peut tirer deux informations de ce résultat :

D'une part, les données d'expression analysées dans cette étude montre bien que les tissus sains n'expriment pas CT83, y compris la glande mammaire normale. Si on revient sur le niveau de méthylation des CpG de CT83 dans les échantillons de tumeurs basales et de sein normal du TCGA (panneau de droite), on observe également cet étalement des valeurs de méthylation dans des échantillons pourtant négatifs pour l'expression de CT83 (en noir), mais pouvant présenter une hypométhylation importante. Il semble donc exister des situations où l'hypométhylation du promoteur de CT83 ne suffit pas à son expression.

D'autre part, CT83 se situant sur le chromosome X, il est possible que l'inactivation d'un des deux X joue un rôle dans le contrôle de l'expression de CT83. Il faudrait évaluer le caractère mono- ou bi-allélique de l'activation de CT83 dans les cancers, et de son hypométhylation dans les tissus sains. On peut imaginer que les mécanismes d'inactivation du X influencent cette répression de CT83 dans les tissus somatiques sains, et que d'autres mécanismes épigénétiques s'ajoutent pour réprimer l'expression de l'allèle de CT83 situé sur le chromosome X hypométhylé.

A HORMAD1 and CT83 methylation status in a human cohort (n = 13 866)

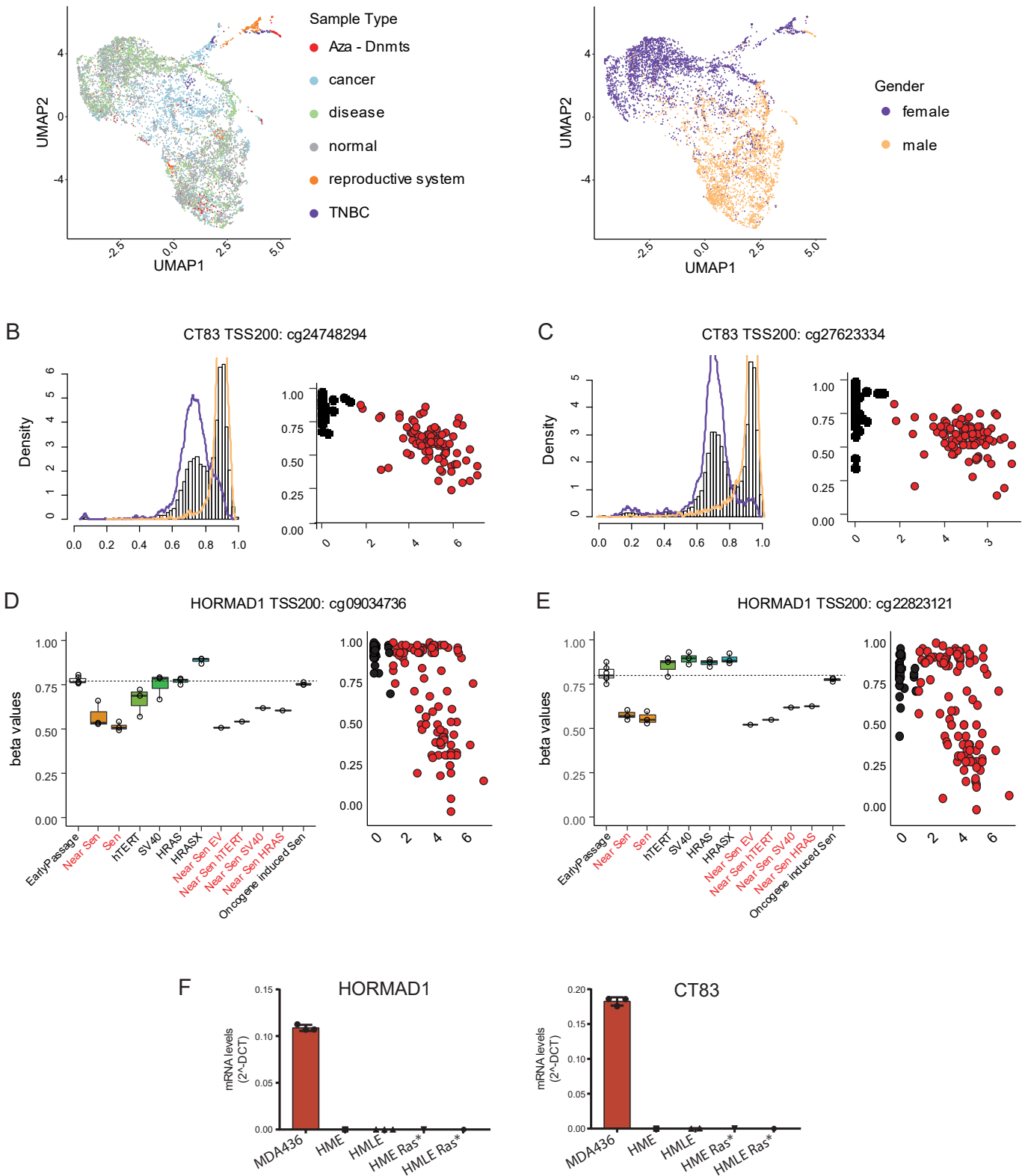


Figure 55 : Piste pour la découverte des mécanismes d'activation d'HORMAD1 et de CT83

A. Analyse multidimensionnelle d'une large cohorte d'échantillons humains (n = 13 866) en fonction du niveau de méthylation de plusieurs CpG situés dans la région promotrice d'HORMAD1 (n = 7) et CT83 (n = 8), évalués par puce 450K. Gauche : annotation des échantillons en fonction de leur origine ou des conditions expérimentales ; Droite : annotation des échantillons en fonction du genre du patient / de la patiente.

B. Distribution du niveau de méthylation d'un CpG situé dans le promoteur de CT83 Gauche: en fonction du genre (violet: féminin, orange: masculin) dans la cohorte d'échantillons humain. Droite : en fonction de l'expression de

Concernant les variations du niveau de méthylation du promoteur d'HORMAD1, nous avons identifié une situation particulière d'hypométhylation de ces CpG (**FIGURE 6D-E**) : il s'agit de la sénescence. Dans cette étude (GSE91069), les auteurs s'intéressent des modifications de la méthylation de l'ADN dans la sénescence et la transformation tumorale. Pour ce faire, ils ont travaillé sur un modèle de fibroblastes humains fraîchement établis, qu'ils ont ensuite transformés selon le schéma classique établi par le laboratoire de Weinberg en 1999 (hTERT + SV40 + HrasV12), ou cultivés jusqu'à obtention d'un état proche de la sénescence (positif pour certains marqueurs de sénescence mais toujours proliférant, apparaissant au bout de 14 divisions cellulaires ou 28 jours de culture environ), ou de sénescence répliquative totale (Xie et al. 2018). En réanalysant les données de cette étude, j'ai pu montrer une hypométhylation significative de certains CpG présents dans le promoteur d'HORMAD1 dans les états proches de la sénescence ou sénescence. De façon intéressante, la transformation seule ne parvient pas à reproduire cet état d'hypométhylation : par exemple la sénescence induite par des oncogènes entraîne une méthylation normale de ces CpG. D'une part, ces résultats confirment les analyses d'expression d'HORMAD1 que nous avons effectués par RTqPCR et en analysant des données publiques de RNAseq dans les lignées dérivées des HMECs : les HMLER, totalement transformées, restent pourtant négatives pour l'expression d'HORMAD1 et de CT83 (**FIGURE 6F**). Ce résultat semble montrer que la transformation seule n'est pas suffisante à l'activation d'HORMAD1 et de CT83. D'autre part, certaines situations physiologiques peuvent conduire à l'hypométhylation du promoteur d'HORMAD1. Il serait intéressant de savoir si ces cellules sénescences 1. expriment HORMAD1 ; 2. sont également les cellules positives pour l'expression de HORMAD1 que l'on a identifié par scRNA-seq dans la glande mammaire saine.

Enfin, j'ai voulu élargir la question de l'activation d'HORMAD1 et de CT83 au delà du cas des cancers du sein. Pour cela, j'ai réalisé une étude pan-cancer basée sur les données du TCGA : j'ai calculé, pour chaque type de tumeur, le pourcentage de tumeurs positives pour HORMAD1 et/ou CT83 (**FIGURE 7**). Ces tumeurs ont été définies comme positives en suivant le même protocole que dans le cas des tumeurs du sein, à partir d'un seuil d'expression défini sur la base du profil d'expression dans les échantillons de tissus normal juxta-tumoral correspondants. Tout d'abord, on observe encore ce caractère non-aléatoire de l'activation d'HORMAD1 et de CT83 : certains types tumoraux sont plus propices que d'autres à ces activations anormales, laissant supposer l'existence de conditions initiales favorisant l'expression d'HORMAD1 et/ou de CT83 au cours de la tumorigenèse.

La co-expression d'HORMAD1 et de CT83 est un événement qui n'est pas exclusif aux cancers du sein : on le retrouve fréquemment dans les tumeurs du poumon (LUAD: 24%, LUSC : 20%), de l'œsophage (23%) et de l'estomac (19%). Cette co-altération est également retrouvée avec une fréquence intermédiaire, similaire à celle observée dans les cancers du sein tout sous-type confondus (11%), dans les cancers du col de l'utérus (10%), de la vessie (9%) et de la tête et du cou (7%). On peut se demander ce que ces types de cancers ont de commun. Les tumeurs hématologiques ou les tumeurs cérébrales n'expriment que très rarement ces deux gènes : les tumeurs positives sont des cancers de types provenant plutôt de la transformation de cellules épithéliales ou squameuses. Cependant ce critère n'est pas suffisant

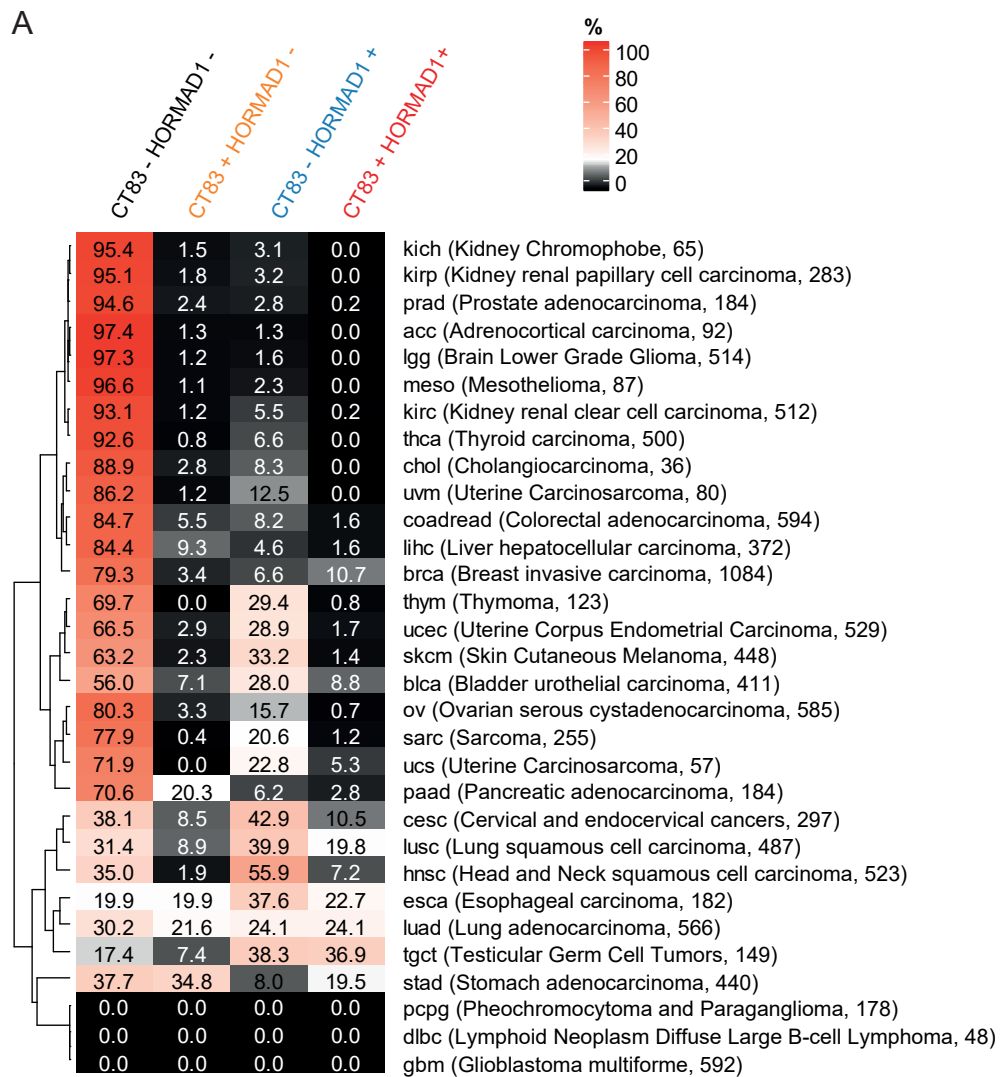


Figure 56 : Influence de la cellule d'origine pour l'activation d'HORMAD1 et de CT83

A. Heatmap représentant le pourcentage de tumeurs exprimant HORMAD1 et / ou CT83, en fonction du type tumorale, au sein de la cohorte du TCGA. L'expression a été évaluée telle qu'elle a été décrite pour la cohorte de tumeurs du sein.

: un contre-exemple peut-être celui des tumeurs du colon, qui se développent également à partir d'un épithélium régionalisé. Les tumeurs du sein triple-négatives sont souvent déficientes pour la recombinaison homologue ; cette altération est également fréquente dans les cancers de l'ovaire, du pancréas, de la prostate, de l'estomac et du poumon (Rodrigues et al. 2016). Là encore, le recouvrement existe mais est imparfait : les tumeurs de l'ovaire et de la prostate expriment rarement HORMAD1 et CT83. Enfin, l'expression de CT83 seul se retrouve dans les cancers majoritairement féminin que sont les cancers de l'utérus (23-30%) et des ovaires (16%), soulignant éventuellement une interférence entre les mécanismes d'inactivation du X et l'activation de CT83 dans ces tumeurs. Une étude pan-cancer plus approfondie, intégrant une description de la nature des cellules d'origine présumées et des altérations oncogénétiques fondatrices (en commençant par les altérations typiques des tumeurs triple-négatives, telles que la déficience de la recombinaison homologue ou l'activation des voies Notch et Myc, Sanchez-Vega et al. 2018), ainsi qu'une estimation du niveau global d'altérations génomiques et épigénétiques, pourrait nous offrir des pistes au sujet des facteurs favorisant l'expression de ces deux gènes C/T.

DISCUSSION



Résumé des résultats

Au cours de ma thèse, nous avons mis en place deux stratégies exploratrices visant à identifier les mécanismes à l'origine de l'activation anormale des gènes Cancer/Testis, et les conséquences de ces expressions ectopiques. Ces deux stratégies reposent sur des approches différentes mais complémentaires : la première utilise plusieurs cribles expérimentaux visant à identifier les interactions entre des acteurs signalétiques et des enzymes épigénétiques contrôlant l'expression de gènes C/T, dans un modèle cellulaire non-tumorigénique. La seconde tire profit de la grande quantité de données disponibles sur les tumeurs du sein, afin de corréler l'activation anormale des gènes C/T et la dérégulation de voies de signalisation dans les cancers, grâce à l'analyse bioinformatique.

Grâce à ces deux stratégies, nous avons étudié plus en détail trois gènes C/T prometteurs, anormalement exprimés dans certains contextes physiopathologiques (**FIGURE 1**). L'activation de ces gènes C/T est non-stochastique et nous avons pu identifier quelques-uns des facteurs impliqués dans la régulation de leur expression, parmi lesquels la voie non canonique du TGF- β pour ADAM12 et l'hypométhylation locale des promoteurs des gènes codant pour *HORMAD1* et *CT83*. De façon intéressante, les mécanismes à l'origine de l'activation de ces gènes C/T sont conservés entre différents types de cancers et entre cellules non-cancéreuses et cancéreuses. L'expression anormale de ces gènes C/T favorise l'acquisition de caractères pro-oncogéniques comme la migration cellulaire ou l'altération de l'homéostasie de la différenciation cellulaire, et pourrait donc favoriser le développement tumoral.

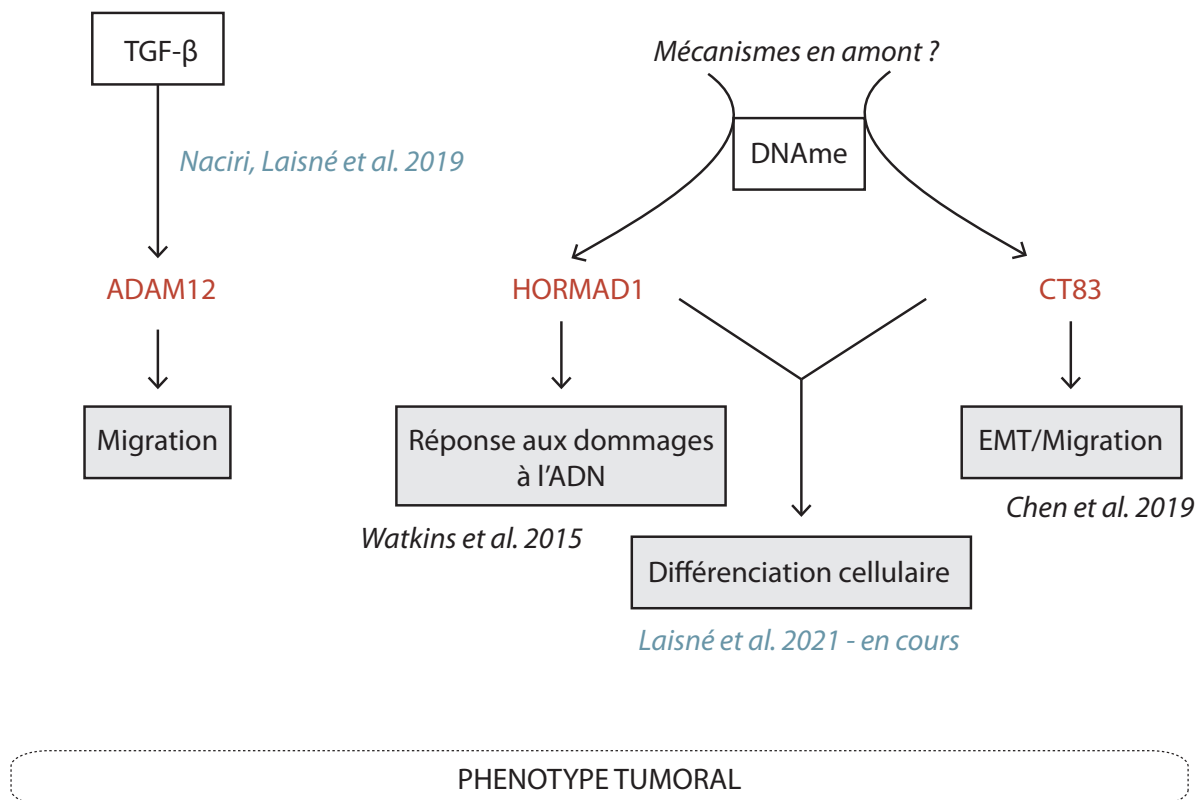


Figure 57 : Activation des gènes C/T et transformation tumorale

Ce schéma récapitule les grands axes dans lesquels sont impliqués les trois gènes C/T que nous avons étudié au cours de ce travail de thèse. DNAme: méthylation de l'ADN

Réflexion autour des stratégies expérimentales et de leurs résultats

Revenant sur ces deux stratégies, il est intéressant de noter que si le premier projet nous a offert des résultats concernant les mécanismes d'activation de certains gènes C/T et en particulier d'ADAM12, le second projet a fourni des résultats riches en hypothèses concernant les associations entre plusieurs gènes C/T et des contextes tumoraux particuliers. Ces deux approches, que nous avons pensées comme réciproques et symétriques, se sont en réalité révélées complémentaires.

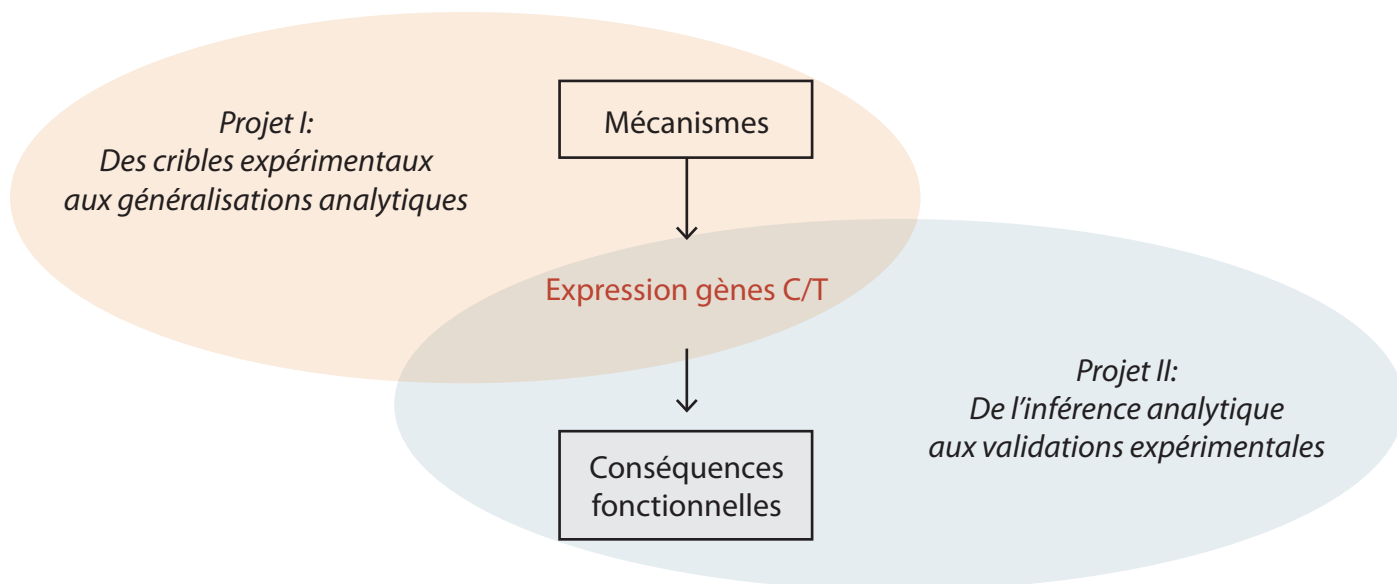


Figure 58 : Bilan des résultats obtenus par nos deux approches expérimentales

Ce schéma présente les trois grandes questions qui s'articulent autour du projet : quels sont les mécanismes à l'origine des expressions anormales des gènes C/T ; quels sont les gènes C/T activés en fonction du contexte ; quelles sont les conséquences fonctionnelles de ces activations anormales.

L'approche expérimentale développée lors du premier projet nous a permis d'organiser la séquence d'événements permettant de moduler l'expression du gène C/T ADAM12. Nous sommes partis d'une situation expérimentale exploratoire («les acteurs signalétiques et épigénétiques testés sont-ils responsables de l'activation hors-contexte des gènes C/T ?») où nous avons obtenu (au moins) un hit, ADAM12 : les cribles réalisés en IMR90 ont identifié la chaîne d'implications causales allant de l'altération des modulateurs TAK1 et KAT2A jusqu'à la surexpression d'ADAM12. Une fois ce mécanisme démontré dans nos modèles cellulaires, une approche bioinformatique restreinte, dirigée par une question précise («l'activation de la voie non-canonique du TGF- β est-elle corrélée à l'expression d'ADAM12 dans les tumeurs du sein, du colon et du poumon ?»), nous a autorisé à supposer ce mécanisme comme généralisable à des situations plus nombreuses.

Notre idée initiale, concernant le second projet sur les cancers du sein, était d'utiliser l'inférence bioinformatique pour découvrir des corrélations solides entre l'activation de gènes C/T spécifiques et la dérégulation de voies signalétiques oncogéniques particulières. Dans cette deuxième étude, notre approche bioinformatique était dirigée par une question assez large («peut-on identifier des gènes C/T anormalement activés dans certains groupes de tumeurs du sein, et dont l'activation pourrait dépendre

d'altérations communes à ces tumeurs ?»). Or, ces analyses nous ont révélé un certain nombre de gènes C/T, 139 exactement, présentant de tels propriétés : au moins 6 d'entre eux se sont révélés associés au sous-type tumoral par exemple. Pour finir ce projet dans le temps imparti, nos approches expérimentales se sont concentrées sur deux gènes C/T prometteurs, *HORMAD1* et *CT83*. Dans le contexte de cette étude, nous avons montré l'avantage qu'offre la co-activation d'*HORMAD1* et *CT83* dans la progression tumorale, et suggère un rôle de ces deux gènes C/T dans l'acquisition des caractéristiques des tumeurs du sein basal-like.

Cette étude pourrait être le début d'un projet plus vaste, visant à décortiquer le rôle éventuel des 137 autres gènes C/T dans la tumorigenèse. Certains de ces gènes seraient particulièrement intéressants à étudier en première approche : par exemple, le gène C/T *DMRTC2*, dont le produit d'expression est une protéine liant l'ADN impliqué dans la régulation transcriptomique des cellules germinales et spécifiquement activé dans un sous-ensemble de cancers du sein HER2-enrichies. Les tumeurs du sein HER2-enrichies sont un groupe plus homogène que les tumeurs basales ; cependant certaines différences sont notables : pour commencer, toutes ne sont pas HER2-amplifiées (Godoy-Ortiz et al. 2019), de plus une proportion significative (45%) d'entre elles sur-expriment également les récepteurs hormonaux (Marchio et al. 2021), enfin ces tumeurs peuvent présenter un mosaïcisme important quant à l'amplification d'*ERBB2* lorsqu'elle est détectée. Ces variations ont d'importantes conséquences sur la prise en charge et le pronostic des patientes. Il serait intéressant d'évaluer l'association éventuelle de *DMRTC2*, activé dans X % des tumeurs du sein, avec un de ces sous-groupes de tumeurs HER2-enrichies, et avec l'amplification d'*ERBB2* à l'échelle de la cellule unique : en effet ce gène C/T présente des caractéristiques intéressantes pour être un biomarqueur efficace. Ceci pourrait faire l'objet d'un nouveau projet dans la continuité du crible réalisé.

Quelles sont les mécanismes d'activation des gènes C/T ?

Pour *ADAM12*, nous avons démontré l'implication de la voie non-canonique du TGF- β dans la régulation de son expression. Cependant, nous nous sommes demandé pourquoi ces deux cribles, visant un nombre considérable de facteurs impliqués dans la régulation de l'expression des gènes, ont uniquement permis l'activation de ce gène C/T parmi les 41 autres.

En effet, le cas d'*ADAM12* se distingue d'autres gènes C/T par plusieurs aspects : tout d'abord, il s'agit d'un gène tissu-enrichi plutôt que tissu-spécifique. Son expression est faible mais détectable dans les IMR90 à l'état basal : la chromatine qui l'entoure se situe dans une conformation plutôt permissive à la transcription. Selon cette hypothèse, et contrairement aux 41 autres gènes C/T évalués dans le crible qui sont probablement réprimés par de multiples couches épigénétiques, *ADAM12* serait plus facilement modulable par un unique acteur épigénétique ou signalétique. En effet, bien qu'*ADAM12* présente un îlot CpG dans son promoteur, celui-ci est faiblement méthylé par RRBS ou méthylArray en IMR90 ; d'autre part son promoteur ne montre pas d'enrichissement particulier dans les marques typiques de l'hétérochromatine que sont H3K9me3 et H3K27me3. Enfin, parler de régulation épigénétique d'*ADAM12*

est peut-être un abus de langage : en effet, notre projet n'a pas véritablement exploré l'existence d'une mémoire moléculaire qui modulerait le niveau d'expression d'ADAM12 après extinction de l'axe TGF-beta / TAK1 / KAT2A. Bien que plusieurs résultats pointent vers un rôle des histones acétyl-transferases dans la régulation d'ADAM12 (Barter et al. 2010 pour HDAC3, KAT2A et SIRT6 dans notre étude), nous n'avons pas montré de modification épigénétique des séquences régulatrices du gène codant pour ADAM12. Des expériences complémentaires seraient nécessaires pour cela, en réalisant par exemple des ChIP-qPCR avec des anticorps reconnaissant H3K9ac ou H3K27ac dans la région promotrice d'ADAM12, avec ou sans stimulation par le TGF-β.

Pour *HORMAD1* et *CT83*, la situation initiale est caractéristique de la majorité des gènes C/T : les deux gènes sont fermement réprimés dans les cellules somatiques saines, et leur expression est strictement restreinte aux cellules germinales mâles. Nous n'avons pas mis à jour la séquence précise d'événements permettant l'activation de *CT83* et *HORMAD1*. Toutefois, nous avons identifié plusieurs facteurs qui jouent un rôle significatif dans leur activation, et qui pris ensemble pourraient conditionner la probabilité d'activation de ces deux gènes.

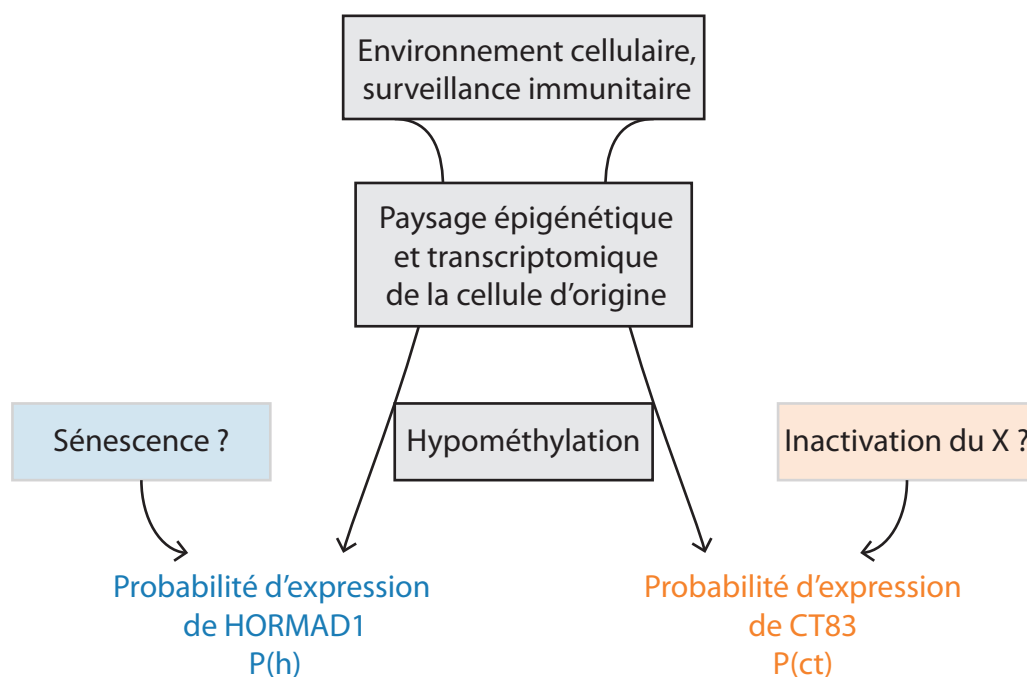


Figure 59 : Facteurs pouvant influencer la probabilité d'activation d'*HORMAD1* et de *CT83*

Certains facteurs sont communs à *HORMAD1* et *CT83* : l'hypométhylation de leurs promoteurs, l'affaiblissement de l'homéostasie tissulaire et de la vigilance immunitaire sont trois facteurs permissifs, qui agissent sur l'activabilité d'*HORMAD1* et de *CT83* dans le contexte tumoral. Ce sont des événements fréquents dans les tumeurs basales, qui sont globalement hypométhylées par rapport aux autres types de tumeurs du sein, et offrent donc un contexte plus favorable à l'activation d'*HORMAD1* et de *CT83*. Ces événements peuvent également se produire dans le contexte de la glande mammaire saine, mais avec une probabilité beaucoup plus faible, ce qui explique la faible fréquence de cellules saines *HORMAD1*-

ou CT83-positives. De plus, ces trois facteurs affectent l'expression d'HORMAD1 et de CT83, ce qui pourrait expliquer la fréquence significative de leur co-activation alors que les deux gènes n'agissent pas l'un sur l'autre et se situent à des loci distincts.

Cependant, ces trois facteurs permissifs sont assez généraux : d'une part, on retrouve dans de nombreux types de cancers une hypométhylation du génome et une permissivité immunologique et tissulaire. Or, nous l'avons vu, tous les types de cancers n'ont pas la même probabilité d'exprimer HORMAD1 et / ou CT83. De plus, ces trois facteurs pourraient influencer de façon équivalente l'activation de tous les gènes tissu-spécifiques régulés par méthylation de l'ADN. Or, tous les gènes C/T dépendant de la méthylation ne sont pas activés dans les tumeurs basales : il y a donc dans l'histoire de ces tumeurs des facteurs qui influencent spécifiquement l'activation d'HORMAD1 et de CT83 en particulier. Qui plus est, dans certains cas ces deux gènes C/T peuvent être exprimés séparément : les mécanismes dirigeant leur activation peuvent donc être différents. Nous avons observé une association entre la sénescence et l'hypométhylation du locus d'HORMAD1, qui n'affecte pas le locus de CT83. Réciproquement, le locus promoteur de CT83 est fréquemment hypométhylé (-25 à -50%) chez les femmes, faisant penser à un phénomène d'échappement allélique de l'inactivation du chromosome X. En revanche, le genre n'a aucune influence sur le statut de méthylation de HORMAD1.

En mettant à jour ces événements qui influencent la probabilité d'activation d'HORMAD1 et de CT83, on pourrait comprendre pourquoi ces gènes sont activés dans certains cancers spécifiquement. Cependant, il ne suffit pas de jouer sur les probabilités d'activation pour observer l'expression stable d'HORMAD1 et de CT83 dans une tumeur : il faut aussi que ces événements soient positivement sélectionnés.

Comment expliquer les fonctions émergentes de la co-activation d'HORMAD1 et de CT83 ?

Comment se fait-il que la co-activation d'HORMAD1 et de CT83 ait un effet très différent, tant au niveau transcriptomique que phénotypique, de l'activation d'un seul de ces gènes ? Cette idée d'émergence suppose une forme de relation entre HORMAD1 et CT83, qui par leur action mutuelle permettrait «le passage d'un système de facteurs interconnectés à un système de facteurs interconnectés autrement», à l'origine de ces nouvelles propriétés. Répondre à cette question exige de saisir les réseaux, de comprendre ce qui se joue entre les liens qui soutiennent le paysage de Waddington et les effets des réverbérations d'un bout à l'autre du circuit. A ce stade, nous disposons de peu d'éléments de preuves, mais nous pouvons en les organisant imaginer une solution possible. Pour ce faire, rappelons les principaux résultats de notre étude.

HORMAD1, s'il est exprimé seul, entraîne une déplétion majeure mais non immédiate des cellules souches dans les HMLE et les HME. Or, dans les lignées de cancers du sein, la surexpression d'HORMAD1 entraîne une accumulation de dommages à l'ADN, probablement due à l'inhibition de la voie de recombinaison homologue via RAD51 (Watkins et al. 2015). Si cette accumulation de dommages

à l'ADN est également observée dans les HME et HMLE, il est possible que ses conséquences soient plus dramatiques pour les cellules souches, entraînant une augmentation de l'apoptose dans cette sous-population spécifiquement (comme semble le suggérer nos observations microscopiques). Ce point pourrait être exploré par cytométrie, en combinant la détection de marqueurs d'apoptose (comme Annexin V / PI) avec les marqueurs de différenciation que nous avons utilisé, afin d'analyser spécifiquement l'apoptose des cellules souches. Pour valider l'implication directe d'HORMAD1 dans ce phénotype, on pourrait faire exprimer une forme inactive d'HORMAD1, présentant par exemple un domaine HORMA tronqué, et évaluer l'impact de ce produit sur la proportion de cellules souches.

Notons que malgré ces observations cytométriques et microscopiques, le RNA-seq ne révèle aucune différence transcriptomique entre la lignée HMLE HORMAD1-positive et la lignée HMLE contrôle. Ceci peut être expliqué par un «effet bulk» : la disparition d'une sous-population cellulaire rare (moins de 3%) ne se traduira probablement pas par des changements transcriptomiques détectables. En revanche, son expansion par un facteur 3 dans le cas de la co-expression de HORMAD1 et CT83 peut produire des effets détectables sur le transcriptome global.

Effectivement, la co-expression d'HORMAD1 et de CT83 entraîne l'émergence surprenante d'un nouveau phénotype. Ce phénotype est fortement enrichi en cellules souches (3 fois plus) et présente des marques caractéristiques de cellules souches anormales (CD24-négatives). Les cellules sont également plus performantes du point de vue de l'EMT, et semblent présenter d'après nos observations microscopiques des cellules de morphologie mésenchymateuse « normales » et non apoptotiques. Surtout, elles arborent une signature transcriptomique nouvelle, caractérisée par la dérégulation de 88 gènes et conservées dans les tumeurs basales CT83- et HORMAD1-positives. Comment comprendre cette émergence de nouveaux caractères ? HORMAD1 et CT83 ne semblent pas avoir d'interaction directe : en effet les expériences d'immunofluorescence montrent qu'ils appartiennent tous deux à des compartiments cellulaires différents ; de plus leur localisation subcellulaire est identique lorsqu'ils sont exprimés seuls ou conjointement.

L'émergence de ces nouveaux effets phénotypiques et transcriptomiques pourrait être le résultat d'une addition des perturbations infligées aux réseaux de régulation cellulaire, permettant soit une réorganisation de l'ensemble des réseaux, soit l'activation d'une forme de potentiel de seuil qui déclencherait la mise en branle de nouvelles réactions telles que la transdifférenciation des progéniteurs luminaux.

Il est également imaginable, par une synergie plus directement additive, que l'effet pro-EMT de CT83 (Chen et al. 2019) induise une plus grande plasticité cellulaire dans les cellules les moins différenciées spécifiquement. Ainsi, CT83 permettrait peut-être (on peut l'imaginer), aux cellules souches de mieux résister aux conséquences de l'expression d'HORMAD1 sur la réponse aux dommages à l'ADN, et entraîne la mise en branle de mécanismes adaptatifs supplémentaires, à l'origine des signatures pro-tumorales détectées dans le RNA-seq des HMLE. Les approches en cellules unique pourraient là encore nous apporter des renseignements supplémentaires.

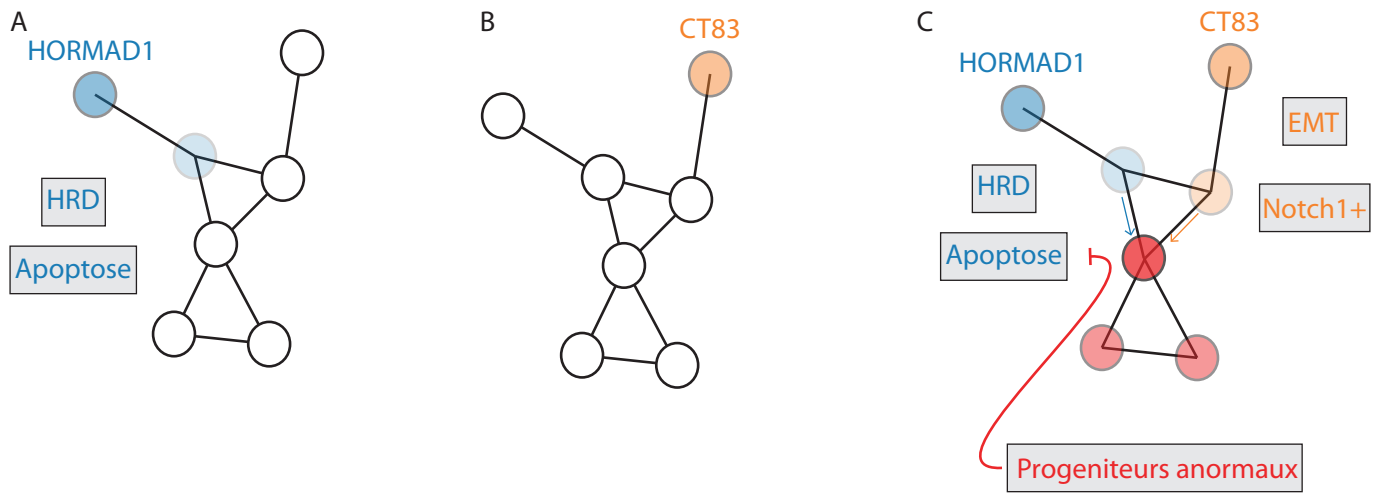


Figure 60 : Modélisation des hypothèses à l'origine de l'émergence de nouvelles propriétés suite à la co-activation d'HORMAD1 et de CT83

A. Activation d'HORMAD1 seul. B. Activation de CT83 seul. C. Co-activation d'HORMAD1 et de CT83

Enfin, en revenant sur ce lien entre activation anormale d'HORMAD1 et réponses aux dommages à l'ADN, on observe une association préférentielle entre des types de cancers déficients pour la recombinaison homologue (ce sont des sous-groupes de cancers de l'ovaire, du poumon, de l'œsophage, de la vessie, tête et cou et de l'estomac, [Knijnenburg et al. 2018](#)) et l'activation d'HORMAD1. Il est possible que l'activation d'HORMAD1 participe directement à l'acquisition du phénotype de *BRCAness*, comme le suggèrent les travaux de Watkins et al. de 2015 dans les cancers du sein. Une autre hypothèse serait que l'effet délétère sur la stabilité du génome induit par HORMAD1 ne soit tolérable pour la cellule qu'en présence d'altérations déjà permissives à une telle instabilité : en quelque sorte, qu'HORMAD1 ait un effet bénéfique ou neutre à condition que les cellules cancéreuses soient à même de supporter un certain niveau de stress génotoxique. Il est également possible que la réponse soit différente en fonction du contexte cellulaire : cela expliquerait pourquoi HORMAD1 semble jouer un rôle dans l'acquisition de la déficience de la recombinaison homologue dans les cancers du sein, alors qu'il serait plutôt un marqueur de proficence de cette même voie dans les cancers du poumon ([Nichols et al. 2018](#)).

A ce stade, ces hypothèses sont purement spéculatives. Cependant, ce modèle pourrait expliquer les différences observées entre la co-expression d'HORMAD1 et de CT83, qui ont un effet sur la différenciation cellulaire, et un traitement par un agent déméthylant tel que la 5-Aza-dC. Ce traitement est à même d'induire l'activation d'HORMAD1 et de CT83 dans les HMLE. Cependant, la 5-Aza-dC ne suffit pas pour reproduire les altérations de différenciation cellulaire observées (données non présentées). Or, la 5-Aza-dC a un effet pléiotropique : de nombreux gènes vont se trouver dérégulés, et les produits d'expressions de ces gènes vont eux aussi impacter de façon massive les réseaux de régulation cellulaire. D'autre part, le mécanisme d'action de la 5-Aza-dC génère un grand nombre de cassures doubles brins, à l'origine d'un stress cellulaire important. Certains travaux ont pu directement montrer qu'un traitement des cellules épithéliales mammaires par des agents épigénétiques tels que la 5-Aza-dC inhibait l'expansion

des progéniteurs luminaux et des cellules souches, via un effet cytostatique (Casay et al. 2018). La 5-Aza-dC aurait alors un effet inhibiteur sur les cellules les moins différenciées des HME / HMLE, masquant les contributions éventuelles d'HORMAD1 et de CT83. Cette interprétation pourrait expliquer la discordance observée entre ces deux expériences, bien qu'HORMAD1 et CT83 soient exprimés dans ces deux conditions.

D'autre part, et toujours en aillant à l'esprit cette notion de seuil, les niveaux d'expression d'HORMAD1 et de CT83 mesurés après un traitement de 48h à la 5-Aza-dC sont très inférieurs à ceux observés dans une lignée de cancer du sein basale (les MDA-MB-436), ainsi qu'aux niveaux obtenus dans nos lignées transgéniques. Pour étudier d'avantage l'influence de la quantité d'HORMAD1 et de CT83 sur la réponse cellulaire, il aurait été intéressant d'utiliser d'autres constructions vectorielles, impliquant des promoteurs inductibles ou plus faibles que celui employé. Sur ce point, le système ADAM12/TAK1/TGF- β présentait l'intérêt de pouvoir manipuler les niveaux d'expression d'ADAM12 dans une gamme de variations compatible avec ce qui peut être observé dans le vivant, car travaillant à partir du promoteur endogène dans son environnement chromatinien naturel.

Quel est la chronologie des événements d'activation des gènes C/T et de transformation ?

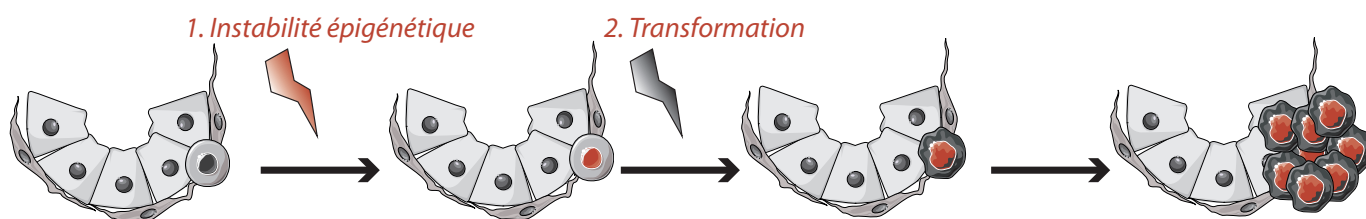
Enfin, concluons cette discussion en abordant l'histoire du développement tumorigénique : Quelle est la chronologie des événements ? L'identité cellulaire doit-elle commencer à s'effriter pour qu'une cellule dérive jusqu'à rompre l'équilibre homéostatique qui justifie sa place au sein de l'organisme ? Ou cette émiettement de l'identité cellulaire est-il une conséquence de la folie cellulaire, incapable d'interpréter correctement les signaux de son environnement et de répondre de façon adéquate à la nécrose, à la suffocation qu'induit sa croissance débridée ? [BLOP BLOP FAUT IL BRIDER CETTE FOLIE VERBEUSE ???^^]

En des termes plus scientifiques, le premier événement est-il l'augmentation de la plasticité épigénétique, permettant la mise en place d'un contexte chromatinien permissif à la fois à l'expression des gènes C/T mais aussi à l'acquisition des caractéristiques fondamentales des cancers ? Ou bien, la transformation survient-elle d'abord, entraînant les défauts chromatiniens responsables de l'activation, plus tardive, des gènes C/T ? D'une part, on a pu montré au cours de cette thèse que la transformation seule ne suffisait ni à engendrer une hypométhylation des promoteurs de CT83 et HORMAD1, ni à provoquer leur activation. Des mécanismes supplémentaires, accompagnant la transformation, doivent nécessairement s'y ajouter. Peut-être est-ce simplement le temps, permettant aux anomalies chromatiniennes d'apparaître et d'être stabilisées, comme le suggèrent les expériences de sénescence sur l'hypométhylation d'HORMAD1. Quoi qu'il en soit, ces expériences suggèrent qu'un certain niveau de déstabilisation chromatinienne est nécessaire pour activer ces gènes C/T, en plus de la dérégulation des voies de signalisation.

D'autre part, nous avons vu que l'expression d'HORMAD1 et de CT83 peut suffire à stimuler certaines

caractéristiques pro-oncogéniques, comme l'EMT ou les anomalies de différenciation. ADAM12, lui, traduirait plutôt l'activation anormale d'un programme d'EMT général, répondant aux signaux de la voie du TGF-beta et englobant son activation. Dans tous les cas, l'expression hors-contexte de gènes C/T peut participer à l'acquisition de propriétés oncogéniques.. Cependant, dans les deux cas, l'ordre des événements entre transformation et anomalies épigénétiques reste à déterminer.

Scénario A: De rares cellules dans le sein non pathologique éprouvent des modifications épigénétiques qui activent les gènes C/T et rendent ces cellules plus susceptibles d'être transformées en cellules tumorales par des mutations oncogéniques.



Scénario B: Après la transformation des cellules à l'origine de la tumeur, certains clones passent par un épisode d'instabilité épigénétique activant les gènes C/T, et seront ensuite positivement sélectionnés, représentant éventuellement la majorité des cellules cancéreuses au diagnostic

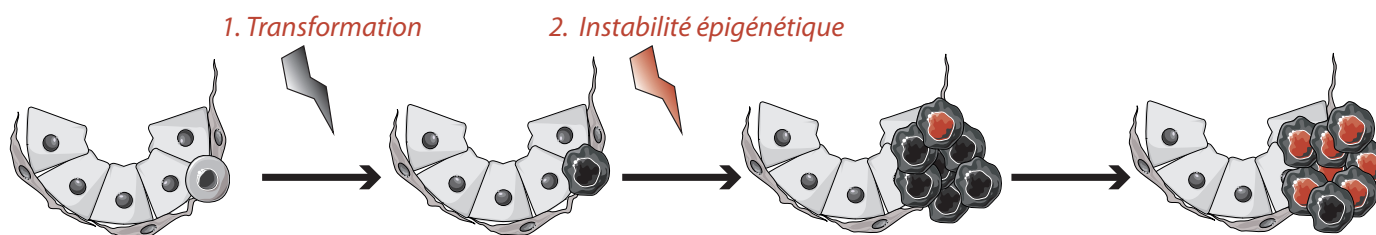


Figure 61 : Deux modèles chronologiques pour l'activation d'HORMAD1 et de CT83

Les cellules HORMAD1- ou CT83-positives sont représentées avec un noyau rouge.

L'activation des gènes C/T est-il un marqueur de progéniteurs anormaux, qui évolueront éventuellement en cellules transformées à l'origine de la majorité des tumeurs basales ? Ou au contraire, l'activation des gènes C/T est-elle un événement tardif, rendu possible par la transformation et favorisé par des forces de sélection positive, jusqu'à représenter éventuellement la majorité des cellules cancéreuses au diagnostic ? La réponse à cette question n'est pas encore claire. La découverte de rares cellules HORMAD1 ou CT83-positives dans des glandes mammaires saines contribue à aviver cette question : ces quelques cellules sont-elles les potentielles cellules d'origine des tumeurs basales, éliminées par des barrières intrinsèques ou par la surveillance environnementale lorsque « tout va bien » ? Sont-elles la marque d'une plasticité épigénétique supérieure à la moyenne, tolérée dans une certaine mesure par l'organisme et démesurément accrue lorsque le cancer trouve à se développer ? Plus encore, sont-elles une première étape accélérant les processus mutagènes et de transdifférenciation qui soutiennent le développement du cancer ? Pour répondre à ces questions, il est crucial de passer du *in vitro* de la culture cellulaire vers des processus expérimentaux plus complexes, tels que la différenciation d'organoides ou l'injection de ces cellules dans des modèles murins humanisés. On pourrait alors étudier l'intégration des cellules de l'épithélium mammaire positives pour HORMAD1 et CT83 au sein d'un arbre mammaire dans une souris, tels que sont capables de former les cellules HME, et d'observer si l'homéostasie de cet

organe en différenciation est d'ores et déjà perturbé par les produits de ces deux gènes. D'un point de vue plus descriptif, il pourrait être très intéressant de trier les rares cellules HORMAD1 ou CT83-positives au sein de la glande mammaire saine, afin de les phénotyper sur la base de l'expression de marqueurs de surface, ou bien de les génotyper pour évaluer l'étendue éventuelle des altérations génotypiques à ce stade, ou encore, par scATACseq par exemple, étudier l'état de plasticité épigénétique de ces cellules en estimant le nombre de régions génomiques différentiellement accessibles.

Du côté de la tumorigenèse, on pourrait réaliser une expérience de transplantation de cellules HMLER contrôles ou positives pour HORMAD1 et CT83 dans des souris *nude*, afin de mesurer la fréquence d'apparition de tumeurs et de caractériser le phénotype de ces tumeurs dans chacun des cas. Enfin, il serait intéressant d'évaluer la part de la plasticité épigénétique que suppose l'activation d'HORMAD1 et de CT83 versus l'action pro-oncogénique avérée par nos expériences de ces protéines. Pour cela, on pourrait imaginer comparer le potentiel tumorigénique de cellules HMLER potentialisées par un bref traitement avec un agent déméthylant par exemple, versus exprimant de façon exogène HORMAD1 et CT83.

Vers des immunothérapies visant les antigènes C/T ?

Ces modèles plus élaborés permettraient de saisir d'un peu plus près la complexité des réseaux de régulation qui régissent l'homéostasie des organismes, et que nous ne pouvons que très imparfaitement reproduire en culture cellulaire simple. En effet, il est certain que les relations entre les cellules cancéreuses et le stroma tumoral, incluant les cellules immunitaires, influencent et conditionnent le développement de la tumeur. Or, les produits d'expression des gènes C/T sont fréquemment immunogéniques : cela a même été précisément démontré pour CT83 dans le cas des cancers de l'ovaire et du sein. L'activation de ces gènes nécessite donc, d'une certaine manière, la complicité du système immunitaire, ce qui pose la question du développement de stratégies thérapeutiques reposant sur la réactivation de l'immunité anti-tumorale, en utilisant les antigènes issus des gènes C/T comme des cibles spécifiques des cellules tumorales.

ANNEXES - Autres contributions scientifiques

Article de recherche

Gupta, N., Yakhou, L., Richard Albert, J., Miura, F., Ferry, L., Kirsh, O., Laisné, M., Yamaguchi, K., Domrane, C., Bonhomme, F., Sarkar, A., Delagrangé, M., Ducos, D., Greenberg, MVC., Cristofari, G., Bultmann, S., Ito, T., Defossez, PA. (2021). A genome-wide knock-out screen for actors of epigenetic silencing reveals new regulators of germline genes and 2-cell like cell state

<https://www.biorxiv.org/content/10.1101/2021.05.03.442415v1>

Revue

Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., and Defossez, P.-A. (2018). Mechanisms of DNA Methyltransferase Recruitment in Mammals. *Genes* 9, 617.

Talks & Posters

2021 : Epigenetics In Cancer Symposium – Cancer Science Institute of Singapore

Talk: Causes & consequences of Cancer/Testis genes activation in Breast Cancer

2019 : Wellcome Genome Campus Advanced Course - Cambridge

Poster : System Biology of Cancer: From Large Datasets to Biological Insight

2018 : The Origin of Cancer – ADELIH, Paris

Poster : Reactivation of epigenetically silenced genes in cancer

2017 : France-Japan Epigenetics Workshop – Université Paris 7 Denis Diderot, Paris

Poster : Targeting kinases to reactivate Cancer/Testis genes in colorectal cancer

Review

Mechanisms of DNA Methyltransferase Recruitment in Mammals

Marthe Laisné¹, Nikhil Gupta¹ , Olivier Kirsh¹ , Sriharsa Pradhan²
and Pierre-Antoine Defossez^{1,*}

¹ Epigenetics and Cell Fate, UMR7216 CNRS, University Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France; marthe.laisne@gmail.com (M.L.); nikhilsspp@gmail.com (N.G.); olivier.kirsh@univ-paris-diderot.fr (O.K.)

² New England Biolabs, 240 County Rd, Ipswich, MA 01938, USA; pradhan@neb.com

* Correspondence: pierre-antoine.defossez@univ-paris-diderot.fr

Received: 16 November 2018; Accepted: 5 December 2018; Published: 10 December 2018



Abstract: DNA methylation is an essential epigenetic mark in mammals. The proper distribution of this mark depends on accurate deposition and maintenance mechanisms, and underpins its functional role. This, in turn, depends on the precise recruitment and activation of de novo and maintenance DNA methyltransferases (DNMTs). In this review, we discuss mechanisms of recruitment of DNMTs by transcription factors and chromatin modifiers—and by RNA—and place these mechanisms in the context of biologically meaningful epigenetic events. We present hypotheses and speculations for future research, and underline the fundamental and practical benefits of better understanding the mechanisms that govern the recruitment of DNMTs.

Keywords: epigenetics; DNA methylation; DNA methyltransferases

1. DNA Methylation: An Essential and Dynamic Epigenetic Mark

The activity of the genome, especially gene expression, is regulated by epigenetic marks. This regulation has to combine two seemingly incompatible objectives: first, the epigenetic marks should be stable enough to contribute to a cell identity that is maintained through the lifetime of the cell, and that is passed on to the daughter cell [1]. Second, the marks have to be flexible enough to allow plasticity [2]. This plasticity can be very local, for instance at the scale of one promoter when a gene is induced by a stimulus. It can also be global, during the large reprogramming events that occur in the zygote after fertilization, or in primordial germ cells as they erase their parental identity to be able to produce gametes [3].

Understanding the stability and the dynamics of epigenetic marks is therefore important to discover the fundamentals of genome activity; and for detection and potential correction of the epigenetic drift that accompanies aging, as well as abnormal epigenetic reprogramming that is an underlying cause of many diseases such as cancers [4]. Finally, from a practical standpoint, understanding the stability and dynamics of chromatin marks is useful to reprogram the epigenome [5]. Again, this can be done on a single-gene scale, or on whole-genome scale to reprogram cells, which constitutes the starting point of regenerative medicine [6].

DNA methylation is a chromatin mark that is essential in mammals. It can rightly be called an “epigenetic” mark, as it has been proven to pass from mother to daughter cells, and sometimes even from one organismal generation to the next [7,8]. The main type of DNA methylation observed in mammals is the methylation of position five of cytosine within a CpG dinucleotide. Non-CpG methylation does occur, for instance in the brain [9,10], but its dynamics and roles are less understood and will not be discussed further here.

In differentiated mammalian cells, about 80% of the CpGs in the genome are methylated; however, there are marked local differences (Figure 1). Intragenic regions and repeated elements are generally methylated. CpG islands, which are associated with the promoters of about two thirds of mammalian genes, are generally unmethylated; conversely methylated promoters are often silenced. Enhancers can also be dynamically methylated, which modifies their ability to recruit transcription factors and activate transcription [11–13]. Finally, the body of actively transcribed genes is methylated, and there is a positive correlation between expression and gene body methylation [14].

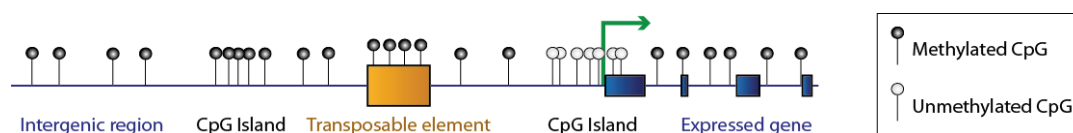


Figure 1. Genome-wide distribution of DNA methylation in mammalian species. DNA methylation occurs symmetrically on CpG sites: to simplify, only one DNA strand is shown on this figure. The mammalian genome has a low frequency of CpGs, but the majority of these are methylated (black lollipops) in intergenic regions, repeated elements and transposable elements, and in gene bodies. Conversely, CpG islands are rich in CpGs and are usually protected from DNA methylation (white lollipops). Unmethylated CpG islands frequently correspond to active promoters (green arrow).

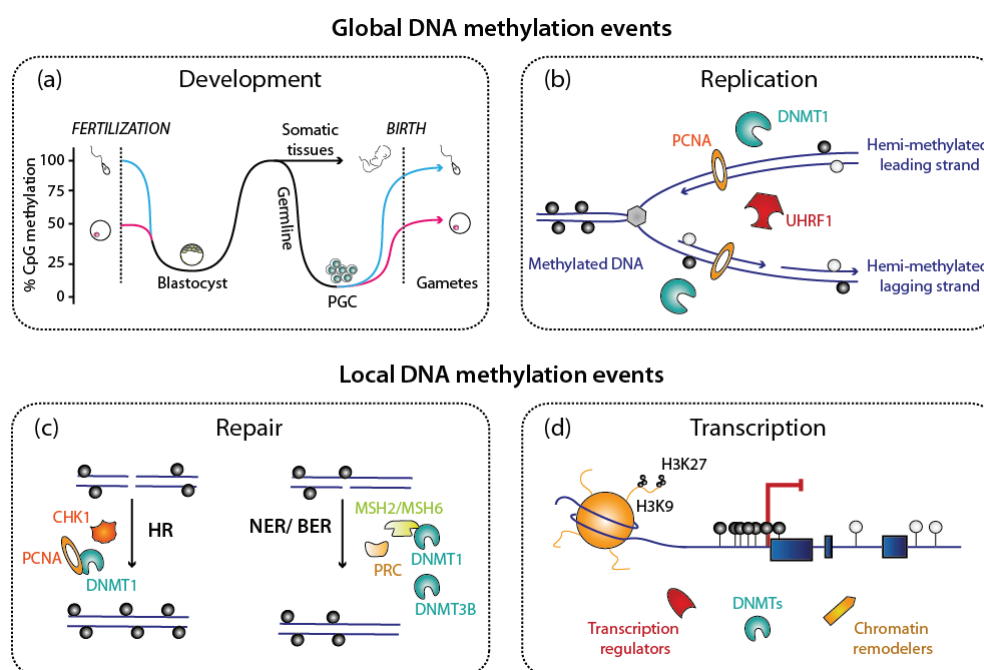


Figure 2. Biological processes involving DNA methylation, DNA methyltransferases (DNMTs) and keys partners. DNA methylation is involved in regulating many cellular processes; for all of them, DNMTs cooperate with several essential actors. (a,b) These DNA methylation events can affect the whole genome, as during the two methylation waves during mammalian life cycle, but also after DNA replicates at each cell cycle. (c,d) DNA methylation events can also be localized, as on the newly synthesized DNA after DNA repair. Selected examples, depicted in this cartoon, will be discussed further in this article. Abbreviations used: PGC: Progenitor germ cells; HR: Homologous Recombination; NER: Nucleotide Excision Repair; BER: Base Excision Repair.

DNA methylation is profoundly remodeled at several steps in the life of a mammalian organism (Figure 2). These steps include early development (Figure 2a), where the zygote after fertilization undergoes massive DNA demethylation, followed by widespread remethylation [15]. Another global demethylation wave occurs in primordial germ cells, as they erase parental imprints,

before re-establishing the DNA methylation pattern found in gametes [3] (Figure 2a). The DNA methylation patterns are also globally challenged at each round of DNA replication (Figure 2b), as the newly synthesized DNA contains only unmethylated cytosines, which then have to be methylated if the parental pattern is to be maintained. Besides these genome-wide transitions, more local events are also observed: when DNA is damaged and repaired, the newly synthesized DNA is initially free of DNA methylation (Figure 2c). Also, local methylation on promoters occurs in the course of transcriptional regulation during development, or in response to a specific stimulus (Figure 2d).

The objective of this review is to summarize and discuss some of the mechanisms that are responsible for the stability and dynamics of DNA methylation, and therefore its functions. For the mark to occur, specific enzymes—the DNA methyltransferases (DNMTs) [16]—have to be recruited to target loci and become catalytically active. In this review, we will describe and discuss recent work on protein- and RNA-mediated recruitment of DNMTs, with a special emphasis on the mammalian enzymes, in the context of diverse functions DNA methylation plays in cellular processes and development.

2. Organization of the DNMTs and Its Functional Consequences

In this section, we will present an overview of the mammalian DNMTs: their domain organization, functions, and potential interacting regions with protein or RNA. The biological role of these interactions will be discussed in subsequent sections.

2.1. Several Non-Redundant Mammalian DNMTs Catalyze CpG Methylation

Cytosine methylation results from the covalent transfer of a methyl group from *S*-adenosyl methionine (SAM) to the carbon C-5 of cytosines to produce 5-methylcytosine (Figure 3). This activity is present in bacterial proteins, such as *M. HhaI*, as part of their restriction/modification systems [17], and iterative searches for mammalian proteins containing a domain similar to the bacterial enzymes led to the identification of the different DNMTs: DNMT1, DNMT2, DNMT3A, DNMT3B, DNMT3C, and DNMT3L (Figures 4 and 5). DNMT3C is present only in rodents, whereas all mammals express the other proteins.

In spite of its similarity to DNA-modifying enzymes, DNMT2 proved to be a tRNA methyltransferase [18,19] and will not be discussed further. DNMT3C is specific to muroids, and was discovered very recently [20]. Our review will for the most part focus on the better-known, catalytically active, enzymes DNMT1, DNMT3A and DNMT3B. We will also discuss DNMT3L: this protein, even though it has no intrinsic catalytic activity, is necessary to stimulate the action of DNMT3A and DNMT3B [21].

Genetic studies in the mouse [22] and other organisms showed that the DNMTs are non-redundant. One reason for their uniqueness is a specific expression pattern [22], but other factors are also involved, as the enzymes have clearly different activities *in vitro* [23]. Remarkably, papers by Riggs [24] as well as Holliday and Pugh [25] proposed as early as 1975 that “maintenance” DNA methylation might be distinguished from “de novo” DNA methylation, and carried out by different proteins. These predictions were validated by experimental work in the subsequent decades: broadly speaking, the maintenance of DNA methylation on hemimethylated CpG sites (generated by DNA replication), is mostly due to DNMT1, which is expressed in all cycling cells; in contrast de novo DNA methylation on both strands of previously unmethylated CpG is mostly carried out by DNMT3A and DNMT3B (Figure 2). There are exceptions to this general division of labor [26,27], but this working model is useful.

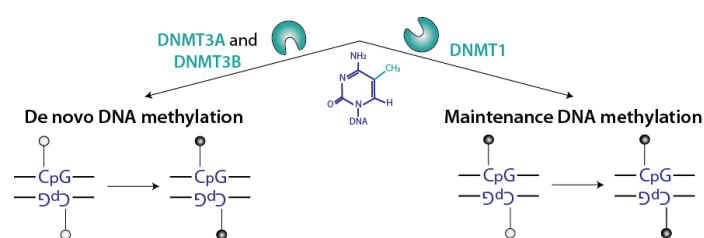


Figure 3. Specificities of the DNMTs DNA methylation can take place in two different contexts, and is processed by two distinct mechanisms: de novo DNA methylation involves fully unmethylated DNA and occurs on both strands, whereas maintenance of DNA methylation involves hemimethylated DNA.

2.2. DNMTs Have Divergent Non-Catalytic Domains

As said above, all vertebrate DNMTs share a conserved region necessary for catalysis; it permits the binding to SAM, and the recognition of DNA. This domain resembles that found in bacterial DNMTs such as *M. HhaI*, emphasizing their shared evolutionary origin (Figures 4 and 5).

Besides this conserved catalytic domain, which is always found at the C-terminus, the different DNMTs contain divergent N-termini that differ in size, as well as in the number and nature of domains they contain. These N-terminal regions contribute to the non-redundant functions of DNMTs.

DNMT1 is the largest of the enzymes (1616 amino acids in humans). Its catalytic domain is separated from the large N-terminal regulatory region by a series of Lys-Gly dipeptide repeats. The N-terminal part harbors different domains: (i) a charge-rich (C-R) domain, which includes a proliferating cell nuclear antigen (PCNA) binding domain (PBD); (ii) an “intrinsically disordered domain”, found only in eutherian mammals [28,29], which contains a nuclear localization sequence (NLS); (iii) a replication foci target sequence (RFTS) domain; (iv) a zinc finger DNA binding domain (CXXC); (v) two bromo-adjacent homology domains (BAH1/2) [30].

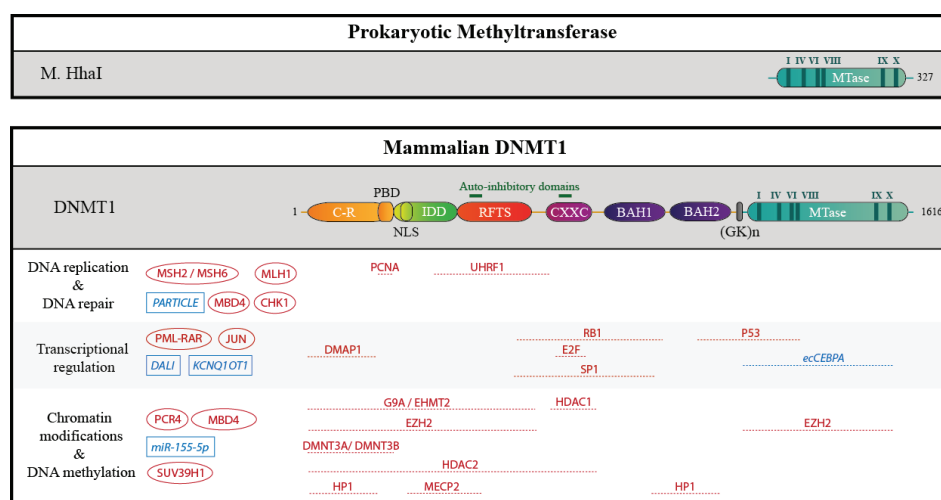


Figure 4. Schematic structure of the prokaryotic methyltransferase *M. HhaI*, compare to DNMT1 and partners. The human DNMT1 contains 1616 amino acids residues. The catalytic methyltransferase domain (MTase, in blue) is very similar to that of the prokaryotic methyltransferase *M. HhaI* and harbors highly conserved motifs (I-X, in dark blue). In addition, DNMT1 harbors a charge-rich (C-R) domain containing the proliferating cell nuclear antigen (PCNA) binding domain (PBD), an intrinsically disordered domain (IDD) with a nuclear localization sequence (NLS), a replication foci target sequence (RFTS), a zinc finger domain (CXXC), and two bromo-adjacent homology domains (BAH 1/2). The catalytic and the regulatory domains are connected by a series of Gly-Lys repeats. Auto-inhibitory domains are highlighted in green. In addition, some interacting proteins and RNAs are represented: if they are known, mapped interaction domains are indicated. Partners with unknown binding sites are shown on the left. Proteins are depicted in red, RNAs in blue.

Enzymes in the DNMT3 family have closely related architectures: DNMT3A and DNMT3B are composed of (i) the N-terminal domain, which is essential for DNA-binding; (ii) a PWWP domain which recognizes H3K36me₃; (iii) an ADD-PHD domain (ATRX-DNMT3B-DNMT3L/plant homeodomain), which binds unmethylated H3K4; and (iv) the catalytic domain (Figure 5). Several isoforms of DNMT3A and DNMT3B, due to alternative promoters or splicing events, have been identified both in human and mouse, and could be involved in different functions. For example, the two major isoforms of DNMT3A shown in Figure 5 have different genomic localizations despite their high similarity [31]. The rodent-specific protein DNMT3C is similar to DNMT3B but lacks the PWWP domain [20]. DNMT3L resembles DNMT3C.

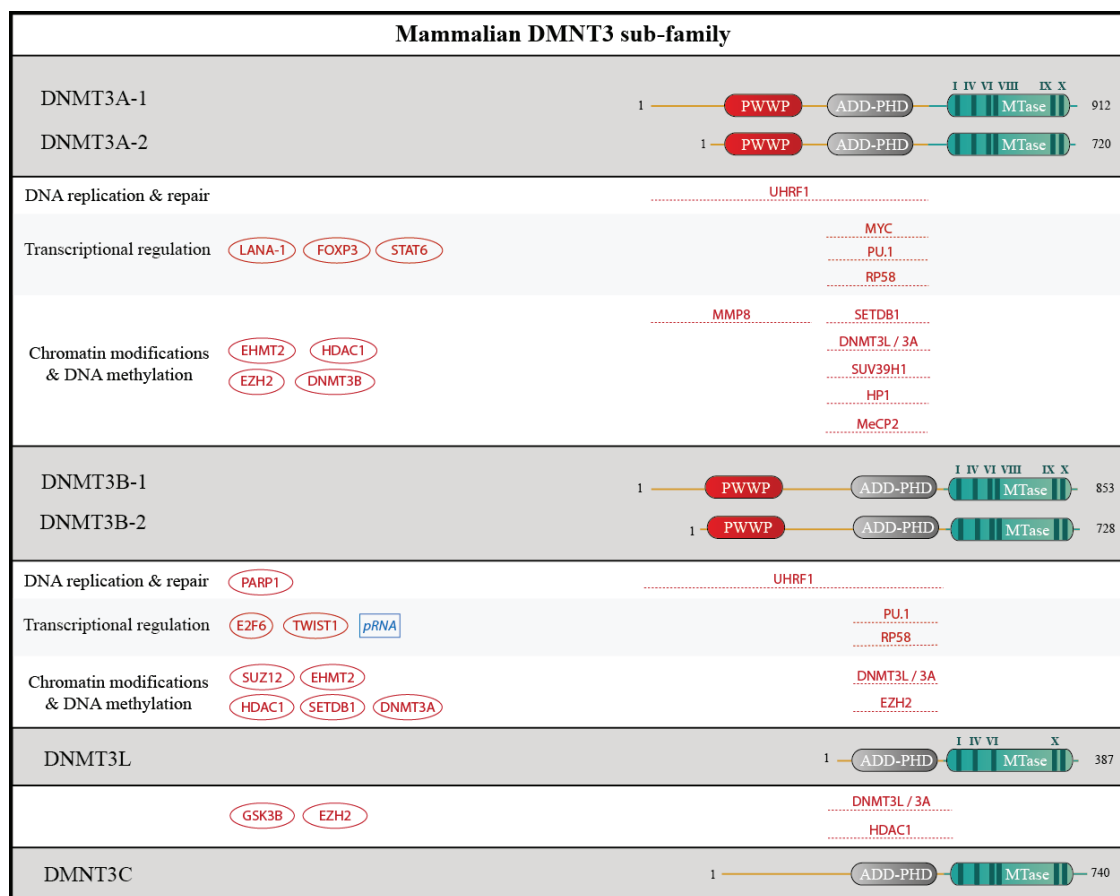


Figure 5. Schematic structure of DNMT3 sub-family and partners. Human DNMT3A, DNMT3B, DNMT3L and the rodent DNMT3C contain 912, 953, 387 and 740 amino acids residues, respectively and for the longest isoforms. For DNMT3A and DNMT3B, transcriptional isoforms due to alternative promoters are shown. The catalytic methyltransferase domain (MTase, in blue) harbors highly conserved motifs (I-X, in dark blue). DNMT3A and DNMT3B comprise a PWWP domain and an ATRX-DNMT3B-DNMT3L and Plant-Homeodomain (ADD-PHD), that is also found in DNMT3L and DNMT3C. Interacting proteins and RNA are depicted as in Figure 4.

2.3. DNMTs Form Complexes

Many of the motifs described above mediate protein-protein interactions, allowing the various DNMTs to participate in multiprotein complexes.

DNMT3A can form hetero-oligomeric complexes: as a heteroduplex with DNMT3L, which increases the processivity of the enzyme, or as a linear heterotetramer with two DNMT3L subunits (at the edges of the tetramer) and two DNMT3A subunits [32]. Moreover, DNMT3A alone

or DNMT3A/DNMT3L complexes can also cooperatively bind DNA, and form large multimeric DNA/protein fibers [33].

Besides homo- or hetero-oligomerization, biochemical approaches such as immunoprecipitation followed by mass spectrometry have revealed that the DNMTs, as other chromatin-modifying enzymes, are part of larger protein complexes [34–36]. Estimating the stoichiometry of these complexes can help reveal which members of the complex are constitutively associated to the DNMTs, and which are minor or transient interactors; this quantitative approach has been historically challenging, but emerging technologies should help improve the situation [37].

2.4. DNMTs Bind Nucleic Acids

The DNMTs contain several domains that can bind nucleic acids, i.e., DNA or RNA. The CXXC domain of DNMT1 is a type of zinc-finger with preferred binding to unmethylated (rather than methylated) CpGs, with an important role in the auto-inhibition mechanism [30]. The catalytic domain also contains a number of basic residues which interact in a sequence-independent manner with the negatively charged DNA backbone [38]. Interestingly, the DNMTs have an affinity for G-quadruplexes [39], and this is biologically relevant [40]. Finally, the PWWP of the DNMT3s has been shown to bind DNA [41].

2.5. DNMTs Are Autoinhibited

Finally, an important functional characteristic of the DNMTs is that they are intramolecularly inhibited, which presumably decreases their off-target activity. Structural and biochemical experiments have that DNMT1 is inhibited by intramolecular interactions between the catalytic site and the RFTD domain or the CXXC domain [5]. DNMT3A is also autoinhibited, albeit by a different domain, the ADD [42,43].

2.6. Functional Consequences for Recruitment Mechanisms

Four important conclusions can be drawn from this overview of the DNMTs. First, the enzymes have a large number of domains, structured or unstructured, with which to establish protein-protein or protein-nucleic acid interactions. Second, some of these domains engage in intramolecular interactions with the catalytic domain and inhibit its activity. Therefore, the recruitment of DNMTs by a protein or RNA interactor may have two separate effects: increasing the local enzyme concentration, but also activating the enzyme at its site of recruitment [42]. Third, the recruitment of a DNMT will not necessarily lead to local methylation: the interaction could in fact break up a catalytically productive complex, or stabilize the auto-inhibited form of the enzyme [44]. Fourth, while DNA methylation is of course the best-known activity of DNMTs, they might possess other important functions that are unrelated to DNA methylation. These functions could be intrinsic, or borne by DNMT interactors: it is clear, for instance, that DNMTs associate with other chromatin-modifying factors, such as histone deacetylases (HDACs) [45]. Therefore, recruiting a DNMT may alter DNA methylation locally, but it could also have other consequences on chromatin.

3. DNMT Recruitment in the Regulation of Chromatin Structure and Gene Expression

DNA methylation is deeply linked to cell identity, as it is a determinant of the cellular transcriptional program [46]. Besides regulating cellular gene expression, DNA methylation is also a key contributor to the transcriptional repression of transposons [21]. These functions of DNA methylation depend on the recruitment of DNMTs with transcription factors and other DNA binding proteins, with histone marks and chromatin modifiers, and with non-coding RNAs. We review these interactions in the subsequent sections, with an emphasis on the most recent data.

3.1. Interaction with Promoter-Bound Transcription Factors

It is well described that DNA methylation status can influence the recruitment of transcriptional regulators [47]. Conversely, transcription factors bound to DNA can also directly recruit the DNA methylation machinery. This was first reported for an oncogenic transcription factor, PML-RAR [48], but the paradigm was rapidly extended to unaltered cellular transcription factors, such as p53 recruiting DNMT1 to silence the *SURVIVIN* promoter [49], and MYC recruiting DNMT3a to silence *p21/CDKN1A* [50]. Since then, many other examples of cellular [46] or viral [47] transcription factors recruiting DNMTs to promoters, via direct interactions, have been discovered. This topic, however, has been discussed in a previous review [48]. Most reported examples concern the recruitment of DNMTs to promoter regions. Interestingly, a previous paper showed that the zinc finger protein ZBTB24, which is found mutated in Immunodeficiency, Centromere instability and Facial anomalies (ICF) syndrome, is likely to recruit DNMT3B to certain gene bodies [49]. This mechanism seems to apply not only to the genes transcribed by RNA Polymerase II (PolII), but also the genes that are targets of PolI [50] or PolIII [51]. Also, while the recruitment of DNMTs seems to generally be accompanied by transcriptional repression, the mechanisms may be varied, at least for DNMT1. In some cases, methylation of DNA by the enzymes seems to be the cause of repression, while in others the enzyme may repress transcription independently of its catalytic activity [51]. This non-catalytic repression seems itself due to protein-protein interactions by which DNMT1 recruits chromatin-modifying enzymes [52].

Besides the relative contributions of catalytic and non-catalytic functions of DNMTs to promoter activity, several general questions await clarification. In a given cell, what fraction of the DNMT molecules is engaged in transcriptional regulation? What are the dynamics of these interactions, and are they regulated by modifications of the transcription factors, the DNMTs, or both? In all but a few examples [53], this regulation is unknown. Do the same mechanisms occur at enhancers [54]? Finally, how can the DNMTs interact with such a large number of unrelated transcriptional regulators? The situation is somewhat reminiscent of the transcription machinery, which can be recruited by many different, apparently unstructured transcriptional activation domains [55]. It would be of interest to determine whether the DNMTs also contain low-complexity regions that function in a similar manner.

3.2. Interaction with Chromatin Modifiers

DNMT1 has different histone-binding partners (histone-methyltransferases, histone deacetylases, but also nucleosome remodelers like SNF2H), mainly recruited through its N-terminal domain, and depicted in Figure 4. An illustrative example is the interaction between DNMT1 and the H3K9 methyltransferase G9a/EHMT2 [56], which helps coordinate DNA and histone methylation after DNA replication. More generally, this crosstalk between the DNA and H3K9 methylation pathways seems fairly prevalent [57], and may be of particular importance at repeated sequences such as centromeres [58].

Less is known about the interactome of DNMT3A and DNMT3B, but some of their chromatin-modifier partners have been identified: for example, DNMT3A interacts with the histone-lysine methyltransferase SETDB1 through its plant homeodomain (PHD) zinc finger and contributes to gene silencing [59]. The relationship between the Polycomb machinery and the DNA methylation machinery is probably more complex than initially thought [57]: while it has been ascertained that EZH2 can in fact recruit DNMT3A to the genome, this recruitment is not sufficient to trigger de novo DNA methylation [60]. The histone-binding protein MPP8 forms a molecular bridge between EHMT1/GLP and DNMT3A, which may help coordinate DNA methylation and H3K9 methylation [61]. The interaction involves the chromodomain of MPP8, which binds a methylated lysine in the N-terminus of DNMT3A [61]. A last example is that the ATRX domain of DNMT3A and the histone acetyltransferase HDAC1 can interact; consequently DNMT3A promotes histone deacetylation near the binding sites of its interactor RP58 [62].

3.3. Integrating Interaction Mechanisms to Dynamically Regulate a Transcriptional Program: The Example of Germline Genes

Genes specifically expressed in the germline permit the formation of gametes; examples of such genes are those necessary for meiosis. Their expression is tightly repressed in somatic cells, and it was recognized early on that many of these genes require DNA methylation for repression [63]. Genetic and molecular experiments showed that the methylation of germline genes is laid by DNMT3B [63], and that this deposition was crucially dependent on the transcription factor E2F6, which recruits DNMT3B to its target sites [64]. This transcription factor/DNMT interaction is one part of a complex web of regulation, as the histone modifying enzyme G9a/EHMT2 is also necessary for DNA methylation to occur on certain germline genes [65]. In addition, E2F6 also takes part in a parallel transcriptional repression mechanism, involving a non-canonical polycomb repressive complex 1 (PRC1) complex [66]. The germline genes therefore provide a clear example of the superposition of DNMT recruitment mechanisms and of repressive pathways [67] (Figure 6).

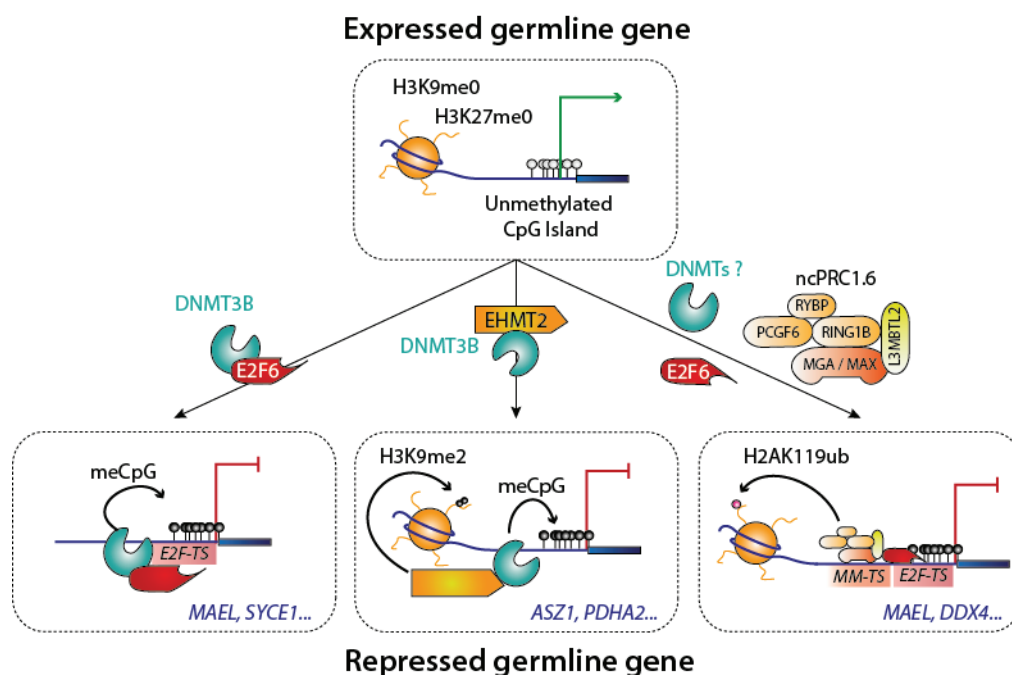


Figure 6. Silencing of the germline genes during implantation. Different complexes are involved in the silencing of germline genes. Three well documented mechanisms are shown, with some example of known targeted germline genes in blue. Abbreviations used: *E2F-TS*: E2F target sequence; *MM-TS*: MGA-MAX target sequence.

3.4. The Role of lncRNAs and miRNAs

Over the past decade, long non-coding RNAs (lncRNAs) have emerged as major regulators of the genome, and they act in part by recruiting nucleoprotein complexes [68]. Mouse and human DNMTs are among the chromatin factors that can be recruited by lncRNAs.

One illustrative instance of this principle was described at the mouse ribosomal DNA (rDNA) locus. There, promoter-associated RNAs (pRNAs), ~200 nucleotides in length, are formed and remain associated with the rDNA promoter via the formation of an RNA:DNA triplex [69]. Fascinatingly, they also associate with DNMT3B, promote its recruitment to the locus, and the inhibition of rDNA transcription [69]. To the best of our knowledge, the region of DNMT3B that is recruited by the pRNA has not been described. However, the recruitment by triplex-forming RNAs has now been shown for DNMT1 as well, as in the case of *PARTICLE*, a lncRNA induced by genotoxic insult [70,71].

Other cases of DNMT cis-recruitment by lncRNA do not seem to involve a triplex-formed RNA. For instance, the *KCNQ1* is a paradigmatic imprinted region, i.e., a region in which the alleles on the paternal and maternal chromosomes have a stable and reciprocal expression pattern. Imprinted genes often produce lncRNA, which participate in the allele-specific regulation of the locus [72]. It was observed that the antisense RNA *KCNQ1OT1*, produced by the paternal allele, interacts with DNMT1, and recruits it to the paternal chromosome, where expression is silenced [73]. The lncRNA-mediated recruitment of DNMTs may also function in trans, as shown with a regulator of neuronal development, *DALI* [74], or within the tumor-suppressive PTEN locus [75].

Although recruitment of DNMTs by RNA is thought to repress the target loci, a recent report challenges this notion [75]. In this paper, it is shown that thousands of loci produce lncRNA which associate with DNMT1. One specific lncRNA, starting upstream of *CEBPA* gene and named *ecCEBPA*, is studied in more detail, and found to adopt a stem-loop structure which binds with high affinity to the catalytic domain of DNMT1. Importantly, this lncRNA protects the *CEBPA* gene against methylation, again illustrating the principle that recruitment of DNMTs does not necessarily equate increased local DNA methylation.

It is noteworthy that microRNAs generated by the ribonuclease DROSHA have been recently shown to bind DNMT1 and decrease its activity [76]. At least one of them, *miR-155-5p*, can apparently act globally and decrease genome methylation [77]. Future work will hopefully reveal whether these mechanisms can also act in recruitment pathways.

3.5. Repression of Transposable Elements: The Role of piRNAs

During its evolution, the mammalian genome has been populated by many transposable elements (TEs). The majority of these sequences are now inactive due to accumulated mutations; however, a small number of copies still retain the potential to transpose. It is now accepted that these mobile elements have a positive role as “evolutionary drivers” [78–80]. Nevertheless, transposition is potentially harmful, and cells have evolved complex mechanisms to tightly control transposons, in part at the transcriptional level [81].

The transcriptional repression of transposons depends heavily on DNA methylation [21], and uses several protein-mediated DNMT recruitment mechanisms. For instance, repression of transposons by TRIM28 leads to DNA methylation [82], although the mechanistic details are lacking. But an original RNA-dependent recruitment pathway is also involved. Indeed, one of the most crucial mechanisms to silence the TEs during male gametogenesis is the PIWI/piRNA pathway in which the piRNAs (small RNAs of 25–32 nt) combine with the PIWI proteins from the Argonaute family to form piRISC complexes. These are directed in a RNA-directed manner to initiate the repression of TEs by recruiting histone modifiers and DNMTs.

PIWI proteins have a defined expression window, among them, MIWI2 expression coincides with the de-novo methylation wave in PGC and also has a nuclear localization, while another critical PIWI protein MILI has a broader expression window and is exclusively cytoplasmic. MIWI2 is proposed to recruit DNMTs, however a detailed mechanism for the recruitment of DNMTs is yet to be elucidated. Further, loss of function studies for MILI and MIWI2 indicate that they may have some non-overlapping roles in DNA methylation of TEs, thus highlighting unknown mechanisms for the recruitment of DNMTs at TEs [81,83,84]. Further, other studies postulate the role of piRNA mediated DNA methylation well beyond the TEs, in the regulation of mRNA transcripts in both somatic and germ cells, imprinted DMR locus and oncogenes involved in cancer [85–87].

4. Maintenance of DNA Methylation during DNA Replication

A very well described role of DNMT1 is to carry out maintenance DNA methylation following DNA replication. This is mediated by a number of well-characterized interactions, which are described in the following paragraphs.

4.1. Interaction with the DNA Replication Machinery

As noted above, it was realized early on that DNA replication would reset the genome to a hemimethylated state, and that if the marks were to be maintained stably through cell divisions, a maintenance mechanism must exist. It was therefore a momentous advance when DNMT1 was shown to directly contact PCNA, providing a direct molecular link between DNA replication and DNA methylation [88]. The exact interacting region was found by pull-down assays against different DNMT1 fragments [89]. It occurs through a conserved PCNA-interacting protein (PIP) motif, a motif which is found in many of the proteins that interact with PCNA and is usually located in intrinsically disordered regions of the proteins [90]. Interestingly, the inactivation of this motif clearly prevents the interaction of DNMT1 with PCNA, and impedes the recruitment of DNMT1 to replication foci in early and mid S phase, but it does not affect the steady-state level of DNA methylation in mouse ES cells [91]. In other words, the direct interaction with PCNA seems to facilitate the recruitment of DNMT1 to replication sites, but other mechanisms can compensate if this interaction is not permitted. We will touch on some of these mechanisms in the following sections.

4.2. Role of UHRF1 and Recognition of Modified Histones

Besides DNMT1, another protein is known to be critically required for DNA methylation maintenance: UHRF1 [92,93]. Initial models suggested that UHRF1 promotes DNA methylation maintenance by directly recruiting DNMT1 through the RFTS domain [94] and then activating the enzyme [95,96], but more recent work suggest that less direct mechanisms are also involved. In particular, UHRF1 is a ubiquitin ligase that can modify histones, and ubiquitinated histones can be bound by the RFTS domain of DNMT1 [97,98]. This important topic is covered in more detail elsewhere in this issue.

4.3. Unresolved Questions

There is no doubt that DNA methylation maintenance is, somehow, coupled to DNA replication. However, as new discoveries are made, it emerges that the simple and elegant model first reported—a direct interaction of DNMT1 with PCNA—coexists with others. Several important questions are still unanswered. The kinetics with which DNA methylation is re-established is a matter of controversy [99,100]. Are lncRNAs involved in DNA methylation maintenance? How much do the “de novo” methyltransferases contribute to maintenance activity, and how [26]? Are the mechanisms of DNA methylation maintenance identical on leading and lagging strand [101]? The answer to these questions is directly linked to the identification of mechanisms recruiting the respective enzymes to their targets.

5. Restoration of DNA Methylation after DNA Damage

After DNA damage, different types of DNA repair can take place: Nucleotide Excision Repair (NER), Base Excision Repair (BER), Non-homologous End Joining (NHEJ) and Homologous Recombination (HR). The choice depends on the type of damage (for instance, presence of adducts versus presence of a double-strand break), and on the position within the cell cycle, which determines whether a sister chromatid is present to serve as a template for repair. Some types of repair, such as HR, entail the resection and resynthesis of large sections of DNA (up to several kilobases in mammals [102]). Therefore, it is expected that the DNA methylation machinery should be recruited to sites of HR to permit the re-establishment of the mark on the newly re-synthesized DNA, and this prediction has indeed been verified experimentally for DNMT1, both by microscopy [103,104] and by biochemical assays [105,106]. It is likely that one mechanism of recruitment is the direct interaction with PCNA, as the kinetics of recruitment of PCNA and DNMT1 are similar, and a region containing the interaction motif is necessary for the recruitment to occur [103]. Nevertheless, it is possible that other recruitment mechanisms also take place; in fact direct interaction with the DNA damage response protein CHK1

has been shown, but has not been worked out in mechanistic detail [107]. Yet other mechanisms may exist: for instance, UHRF1 has been proposed to directly recognize certain types of damaged DNA [108,109], and it would be of interest to determine whether this permits the recruitment of DNMT1 in parallel to the PCNA-driven pathway.

A particular situation leading to DNA damage is oxidative stress [110]. Oxidized bases such as 8-oxo-G are repaired in large part by BER and NER, but not HR [111]. An interesting report showed that oxidative stress led to the formation of large complexes containing DNMT1, DNMT3B, and Polycomb proteins, which were then addressed to previously unmethylated CpG islands [112]. Follow-up work recently clarified the mechanism, which depends on the mismatch repair proteins MSH2/MSH6 [113,114], but not on PCNA. While the interaction of DNMT1 with MSH2/MSH6 is clearly stimulated by oxidative stress, its mechanistic underpinning remains to be precised: what are the domains of the proteins involved? Why is the interaction stimulated by oxidation? These results echo earlier findings showing an interaction of DNMT1 with the mismatch repair protein MLH1 [104].

To summarize, DNMTs are recruited to chromatin after several types of DNA damage, double-strand breaks and oxidative damage have both been proved directly, and it would be interesting to assess whether other lesions, such as single-strand breaks or pyrimidine dimers also have the same effect. One mechanism that is unambiguously involved is the interaction of DNMT1 with PCNA. Interactions with mismatch repair proteins also appear important, but their molecular basis is unclear. Needless to say, the future may reveal yet other modes of DNMT recruitment after DNA damage.

6. Conclusions and Perspectives

6.1. Conceptual Advances in the Roles of DNMTs

The development of high-throughput sequencing technologies has revolutionized our ability to map DNA methylation, to identify genomic loci bound by DNMTs, and to identify RNAs associated with DNMTs. In parallel, advances in mass spectrometry have also made it much easier to detect and quantify protein complexes, while new methods in microscopy give us insight into their location and dynamics. The combination of these methods has allowed the community to arrive at the cumulative knowledge presented in this review. Much of the evidence has cemented early insight that DNA methylation is a critical epigenetic mechanism, which contributes to cell identity by regulating transcriptional programs, by ensuring the proper chromatin composition on key chromosome elements such as centromeres, and by repressing repeated elements. Another important lesson is that DNA methylation is part of a complex mesh of chromatin regulation mechanisms, which includes non-coding RNAs, histone- and nucleosome-modifiers, and DNA demethylation activities [15]. An important issue, not discussed here, is how DNA methylation is coordinated to DNA demethylation to achieve the final patterns seen in cells [115].

6.2. Targeting the DNMTs for Epigenome Editing

Artificially recruiting DNMTs to a locus of interest has long been considered as potentially useful to “edit the epigenome”, for example to turn off the expression of oncogenes in tumor cells, or to remodel the genome of stem cells [116]. The idea has been made much easier to implement with the development of Cas9-based platforms [5,117]. The results described in this review have practical applications. First, they show that non-catalytic activities of DNMTs can be critical for repression and may have to be maintained for a Cas9 fusion to work efficiently. Second, they underline that increasing the local concentration of DNMTs by recruitment does not always translate into increased DNA methylation, as self-inhibitory mechanisms have to be overcome. Third, they show that RNA can be explored as a way to recruit DNMTs. Again, the recruitment can either lead to local DNA methylation and repression [85] or, conversely, to inhibition of the enzyme and local protection from its activity [118].

6.3. Consequences for Disease and Treatment

Most of the mechanisms we have described occur during the development and life of a healthy organism. We have already mentioned, however, that some of these mechanisms can be subverted by viral or oncogenic proteins. Understanding these events molecularly may help target them pharmaceutically to fight infections and cancers. More generally, the DNA methylation patterns drift during human aging, and this may contribute to the increase in cancer risk with age [119]. It will be of great interest in the future to determine if the DNMT recruitment mechanisms that have been identified go awry in aging cells, and whether this can be prevented or reversed pharmaceutically.

Author Contributions: All authors contributed to the writing and correction of the review.

Funding: Work in the lab of P.A.D. was supported by Association pour la Recherche contre le Cancer (ARC2014), by Agence Nationale de la Recherche (ANR-15-CE12-0012-01 and ANR-11-LABX-0071 under ANR-11-IDEX-0005-01), and by Institut National du Cancer (INCa PLBio 2015-1-PLBio-01-DR A-1).

Acknowledgments: We are grateful to members of the Defossez lab for useful discussions. We apologize to the authors of relevant primary papers which could not be cited because of space constraints.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yadav, T.; Quivy, J.-P.; Almouzni, G. Chromatin plasticity: A versatile landscape that underlies cell fate and identity. *Science* **2018**, *361*, 1332–1336. [[CrossRef](#)] [[PubMed](#)]
2. Atlasi, Y.; Stunnenberg, H.G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **2017**, *18*, 643–658. [[CrossRef](#)]
3. Nashun, B.; Hill, P.W.S.; Hajkova, P. Reprogramming of cell fate: Epigenetic memory and the erasure of memories past. *EMBO J.* **2015**, *34*, 1296–1308. [[CrossRef](#)] [[PubMed](#)]
4. Dawson, M.A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* **2017**, *355*, 1147–1152. [[CrossRef](#)] [[PubMed](#)]
5. Kungulovski, G.; Jeltsch, A. Epigenome editing: State of the art, concepts, and perspectives. *Trends Genet. TIG* **2016**, *32*, 101–113. [[CrossRef](#)] [[PubMed](#)]
6. Takahashi, K.; Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 183–193. [[CrossRef](#)] [[PubMed](#)]
7. Schübeler, D. Function and information content of DNA methylation. *Nature* **2015**, *517*, 321–326. [[CrossRef](#)]
8. Luo, C.; Hajkova, P.; Ecker, J.R. Dynamic DNA methylation: In the right place at the right time. *Science* **2018**, *361*, 1336–1340. [[CrossRef](#)]
9. Lister, R.; Mukamel, E.A.; Nery, J.R.; Urich, M.; Puddifoot, C.A.; Johnson, N.D.; Lucero, J.; Huang, Y.; Dwork, A.J.; Schultz, M.D.; et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **2013**, *341*, 1237905. [[CrossRef](#)]
10. Kinde, B.; Gabel, H.W.; Gilbert, C.S.; Griffith, E.C.; Greenberg, M.E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6800–6806. [[CrossRef](#)]
11. Yao, L.; Shen, H.; Laird, P.W.; Farnham, P.J.; Berman, B.P. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* **2015**, *16*, 105. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, Y.; Zhang, D.; Li, Q.; Liang, J.; Sun, L.; Yi, X.; Chen, Z.; Yan, R.; Xie, G.; Li, W.; et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nat. Genet.* **2016**, *48*, 1003–1013. [[CrossRef](#)] [[PubMed](#)]
13. Fleischer, T.; Tekpli, X.; Mathelier, A.; Wang, S.; Nebdal, D.; Dhakal, H.P.; Sahlberg, K.K.; Schlichting, E.; Oslo Breast Cancer Research Consortium (OSBREAC); Børresen-Dale, A.-L.; et al. DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat. Commun.* **2017**, *8*, 1379. [[CrossRef](#)] [[PubMed](#)]
14. Yang, X.; Han, H.; De Carvalho, D.D.; Lay, F.D.; Jones, P.A.; Liang, G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **2014**, *26*, 577–590. [[CrossRef](#)] [[PubMed](#)]
15. Iurlaro, M.; von Meyenn, F.; Reik, W. DNA methylation homeostasis in human and mouse development. *Curr. Opin. Genet. Dev.* **2017**, *43*, 101–109. [[CrossRef](#)] [[PubMed](#)]

16. Gowher, H.; Jeltsch, A. Mammalian DNA methyltransferases: New discoveries and open questions. *Biochem. Soc. Trans.* **2018**, *46*, 1191–1202. [[CrossRef](#)] [[PubMed](#)]
17. Blow, M.J.; Clark, T.A.; Daum, C.G.; Deutschbauer, A.M.; Fomenkov, A.; Fries, R.; Froula, J.; Kang, D.D.; Malmstrom, R.R.; Morgan, R.D.; et al. The epigenomic landscape of prokaryotes. *PLoS Genet.* **2016**, *12*, e1005854. [[CrossRef](#)]
18. Goll, M.G.; Kirpekar, F.; Maggert, K.A.; Yoder, J.A.; Hsieh, C.-L.; Zhang, X.; Golic, K.G.; Jacobsen, S.E.; Bestor, T.H. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* **2006**, *311*, 395–398. [[CrossRef](#)]
19. Defossez, P.-A. Ceci n'est pas une DNMT: Recently discovered functions of DNMT2 and their relation to methyltransferase activity (Comment on DOI 10.1002/bies.201300088). *BioEssays* **2013**, *35*, 1024. [[CrossRef](#)]
20. Barau, J.; Teissandier, A.; Zamudio, N.; Roy, S.; Nalesso, V.; Héroult, Y.; Guillou, F.; Bourc'his, D. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **2016**, *354*, 909–912. [[CrossRef](#)]
21. Edwards, J.R.; Yarychivska, O.; Boulard, M.; Bestor, T.H. DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* **2017**, *10*, 23. [[CrossRef](#)] [[PubMed](#)]
22. Dan, J.; Chen, T. Genetic studies on mammalian dna methyltransferases. *Adv. Exp. Med. Biol.* **2016**, *945*, 123–150. [[CrossRef](#)] [[PubMed](#)]
23. Ambrosi, C.; Manzo, M.; Baubec, T. Dynamics and context-dependent roles of DNA methylation. *J. Mol. Biol.* **2017**, *429*, 1459–1475. [[CrossRef](#)] [[PubMed](#)]
24. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **1975**, *14*, 9–25. [[CrossRef](#)] [[PubMed](#)]
25. Holliday, R.; Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **1975**, *187*, 226–232. [[CrossRef](#)] [[PubMed](#)]
26. Walton, E.L.; Francastel, C.; Velasco, G. Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics* **2011**, *6*, 1373–1377. [[CrossRef](#)]
27. Elliott, E.N.; Sheaffer, K.L.; Kaestner, K.H. The “de novo” DNA methyltransferase Dnmt3b compensates the Dnmt1-deficient intestinal epithelium. *eLife* **2016**, *5*. [[CrossRef](#)]
28. Shaffer, B.; McGraw, S.; Xiao, S.C.; Chan, D.; Trasler, J.; Chaillet, J.R. The DNMT1 intrinsically disordered domain regulates genomic methylation during development. *Genetics* **2015**, *199*, 533–541. [[CrossRef](#)] [[PubMed](#)]
29. Borowczyk, E.; Mohan, K.N.; D'Aiuto, L.; Cirio, M.C.; Chaillet, J.R. Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20806–20811. [[CrossRef](#)]
30. Song, J.; Rechkoblit, O.; Bestor, T.H.; Patel, D.J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **2011**, *331*, 1036–1040. [[CrossRef](#)]
31. Manzo, M.; Wirz, J.; Ambrosi, C.; Villaseñor, R.; Roschitzki, B.; Baubec, T. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* **2017**, *36*, 3421–3434. [[CrossRef](#)] [[PubMed](#)]
32. Jia, D.; Jurkowska, R.Z.; Zhang, X.; Jeltsch, A.; Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **2007**, *449*, 248–251. [[CrossRef](#)] [[PubMed](#)]
33. Jurkowska, R.Z.; Anspach, N.; Urbanke, C.; Jia, D.; Reinhardt, R.; Nellen, W.; Cheng, X.; Jeltsch, A. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res.* **2008**, *36*, 6656–6663. [[CrossRef](#)] [[PubMed](#)]
34. Wan, C.; Borgeson, B.; Phanse, S.; Tu, F.; Drew, K.; Clark, G.; Xiong, X.; Kagan, O.; Kwan, J.; Bezginov, A.; et al. Panorama of ancient metazoan macromolecular complexes. *Nature* **2015**, *525*, 339–344. [[CrossRef](#)]
35. Huttlin, E.L.; Bruckner, R.J.; Paulo, J.A.; Cannon, J.R.; Ting, L.; Baltier, K.; Colby, G.; Gebreab, F.; Gygi, M.P.; Parzen, H.; et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **2017**, *545*, 505–509. [[CrossRef](#)] [[PubMed](#)]
36. Ponnaluri, V.K.C.; Estève, P.-O.; Ruse, C.I.; Pradhan, S. S-adenosylhomocysteine hydrolase participates in DNA methylation inheritance. *J. Mol. Biol.* **2018**, *430*, 2051–2065. [[CrossRef](#)] [[PubMed](#)]
37. Smits, A.H.; Vermeulen, M. characterizing protein-protein interactions using mass spectrometry: Challenges and opportunities. *Trends Biotechnol.* **2016**, *34*, 825–834. [[CrossRef](#)]

38. Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D.J. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **2012**, *335*, 709–712. [[CrossRef](#)]
39. Cree, S.L.; Fredericks, R.; Miller, A.; Pearce, F.G.; Filichev, V.; Fee, C.; Kennedy, M.A. DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett.* **2016**, *590*, 2870–2883. [[CrossRef](#)]
40. Mao, S.-Q.; Ghanbarian, A.T.; Spiegel, J.; Martínez Cuesta, S.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hänsel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **2018**, *25*, 951–957. [[CrossRef](#)]
41. Qiu, C.; Sawada, K.; Zhang, X.; Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* **2002**, *9*, 217–224. [[CrossRef](#)] [[PubMed](#)]
42. Jeltsch, A.; Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases—A new regulatory paradigm. *Nucleic Acids Res.* **2016**, *44*, 8556–8575. [[CrossRef](#)] [[PubMed](#)]
43. Guo, X.; Wang, L.; Li, J.; Ding, Z.; Xiao, J.; Yin, X.; He, S.; Shi, P.; Dong, L.; Li, G.; et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **2015**, *517*, 640–644. [[CrossRef](#)] [[PubMed](#)]
44. Rajavelu, A.; Lungu, C.; Emperle, M.; Dukatz, M.; Bröhm, A.; Broche, J.; Hanelt, I.; Parsa, E.; Schiffers, S.; Karnik, R.; et al. Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic Acids Res.* **2018**, *46*, 9044–9056. [[CrossRef](#)] [[PubMed](#)]
45. Jones, P.A.; Issa, J.-P.J.; Baylin, S. Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.* **2016**, *17*, 630–641. [[CrossRef](#)] [[PubMed](#)]
46. Bogdanović, O.; Lister, R. DNA methylation and the preservation of cell identity. *Curr. Opin. Genet. Dev.* **2017**, *46*, 9–14. [[CrossRef](#)]
47. Yin, Y.; Morgunova, E.; Jolma, A.; Kaasinen, E.; Sahu, B.; Khund-Sayeed, S.; Das, P.K.; Kivioja, T.; Dave, K.; Zhong, F.; et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **2017**, *356*. [[CrossRef](#)]
48. Di Croce, L.; Raker, V.A.; Corsaro, M.; Fazi, F.; Fanelli, M.; Faretta, M.; Fuks, F.; Lo Coco, F.; Kouzarides, T.; Nervi, C.; et al. Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* **2002**, *295*, 1079–1082. [[CrossRef](#)]
49. Estève, P.-O.; Chin, H.G.; Pradhan, S. Human maintenance DNA (cytosine-5)-methyltransferase and p53 modulate expression of p53-repressed promoters. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1000–1005. [[CrossRef](#)]
50. Brenner, C.; Deplus, R.; Didelot, C.; Loriot, A.; Viré, E.; De Smet, C.; Gutierrez, A.; Danovi, D.; Bernard, D.; Boon, T.; et al. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **2005**, *24*, 336–346. [[CrossRef](#)]
51. Dunican, D.S.; Ruzov, A.; Hackett, J.A.; Meehan, R.R. xDnmt1 regulates transcriptional silencing in pre-MBT *Xenopus* embryos independently of its catalytic function. *Development* **2008**, *135*, 1295–1302. [[CrossRef](#)] [[PubMed](#)]
52. Clements, E.G.; Mohammad, H.P.; Leadem, B.R.; Easwaran, H.; Cai, Y.; Van Neste, L.; Baylin, S.B. DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. *Nucleic Acids Res.* **2012**, *40*, 4334–4346. [[CrossRef](#)] [[PubMed](#)]
53. Kumar, D.; Lassar, A.B. Fibroblast growth factor maintains chondrogenic potential of limb bud mesenchymal cells by modulating DNMT3A recruitment. *Cell Rep.* **2014**, *8*, 1419–1431. [[CrossRef](#)] [[PubMed](#)]
54. Petell, C.J.; Alabdi, L.; He, M.; San Miguel, P.; Rose, R.; Gowher, H. An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic Acids Res.* **2016**, *44*, 7605–7617. [[CrossRef](#)] [[PubMed](#)]
55. Hnisz, D.; Shrinivas, K.; Young, R.A.; Chakraborty, A.K.; Sharp, P.A. A Phase separation model for transcriptional control. *Cell* **2017**, *169*, 13–23. [[CrossRef](#)] [[PubMed](#)]
56. Estève, P.-O.; Chin, H.G.; Smallwood, A.; Feehery, G.R.; Gangisetty, O.; Karpf, A.R.; Carey, M.F.; Pradhan, S. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **2006**, *20*, 3089–3103. [[CrossRef](#)] [[PubMed](#)]
57. Rose, N.R.; Klose, R.J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta* **2014**, *1839*, 1362–1372. [[CrossRef](#)] [[PubMed](#)]
58. Déjardin, J. Switching between epigenetic states at pericentromeric heterochromatin. *Trends Genet. TIG* **2015**, *31*, 661–672. [[CrossRef](#)]

59. Li, H.; Rauch, T.; Chen, Z.-X.; Szabó, P.E.; Riggs, A.D.; Pfeifer, G.P. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J. Biol. Chem.* **2006**, *281*, 19489–19500. [[CrossRef](#)]
60. Rush, M.; Appanah, R.; Lee, S.; Lam, L.L.; Goyal, P.; Lorincz, M.C. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. *Epigenetics* **2009**, *4*, 404–414. [[CrossRef](#)]
61. Chang, Y.; Sun, L.; Kokura, K.; Horton, J.R.; Fukuda, M.; Espejo, A.; Izumi, V.; Koomen, J.M.; Bedford, M.T.; Zhang, X.; et al. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nat. Commun.* **2011**, *2*, 533. [[CrossRef](#)] [[PubMed](#)]
62. Fuks, F.; Burgers, W.A.; Godin, N.; Kasai, M.; Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **2001**, *20*, 2536–2544. [[CrossRef](#)] [[PubMed](#)]
63. Borgel, J.; Guibert, S.; Li, Y.; Chiba, H.; Schübeler, D.; Sasaki, H.; Forné, T.; Weber, M. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **2010**, *42*, 1093–1100. [[CrossRef](#)] [[PubMed](#)]
64. Velasco, G.; Hubé, F.; Rollin, J.; Neuillet, D.; Philippe, C.; Bouzinba-Segard, H.; Galvani, A.; Viegas-Péquignot, E.; Francastel, C. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9281–9286. [[CrossRef](#)] [[PubMed](#)]
65. Auclair, G.; Borgel, J.; Sanz, L.A.; Vallet, J.; Guibert, S.; Dumas, M.; Cavelier, P.; Girardot, M.; Forné, T.; Feil, R.; et al. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res.* **2016**, *26*, 192–202. [[CrossRef](#)] [[PubMed](#)]
66. Endoh, M.; Endo, T.A.; Shinga, J.; Hayashi, K.; Farcas, A.; Ma, K.-W.; Ito, S.; Sharif, J.; Endoh, T.; Onaga, N.; et al. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *eLife* **2017**, *6*. [[CrossRef](#)]
67. Mochizuki, K.; Tachibana, M.; Saitou, M.; Tokitake, Y.; Matsui, Y. Implication of DNA demethylation and bivalent histone modification for selective gene regulation in mouse primordial germ cells. *PLoS ONE* **2012**, *7*, e46036. [[CrossRef](#)]
68. Quinn, J.J.; Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **2016**, *17*, 47–62. [[CrossRef](#)]
69. Schmitz, K.-M.; Mayer, C.; Postepska, A.; Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **2010**, *24*, 2264–2269. [[CrossRef](#)]
70. O’Leary, V.B.; Ovsepiyan, S.V.; Carrascosa, L.G.; Buske, F.A.; Radulovic, V.; Niyazi, M.; Moertl, S.; Trau, M.; Atkinson, M.J.; Anastasov, N. *PARTICLE*, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep.* **2015**, *11*, 474–485. [[CrossRef](#)]
71. O’Leary, V.B.; Hain, S.; Maugg, D.; Smida, J.; Azimzadeh, O.; Tapio, S.; Ovsepiyan, S.V.; Atkinson, M.J. Long non-coding RNA *PARTICLE* bridges histone and DNA methylation. *Sci. Rep.* **2017**, *7*, 1790. [[CrossRef](#)]
72. Barlow, D.P.; Bartolomei, M.S. Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*. [[CrossRef](#)] [[PubMed](#)]
73. Mohammad, F.; Mondal, T.; Guseva, N.; Pandey, G.K.; Kanduri, C. *Kcnq1ot1* noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* **2010**, *137*, 2493–2499. [[CrossRef](#)] [[PubMed](#)]
74. Chalei, V.; Sansom, S.N.; Kong, L.; Lee, S.; Montiel, J.F.; Vance, K.W.; Ponting, C.P. The long non-coding RNA *Dali* is an epigenetic regulator of neural differentiation. *eLife* **2014**, *3*, e04530. [[CrossRef](#)]
75. Johnsson, P.; Ackley, A.; Vidarsdottir, L.; Lui, W.-O.; Corcoran, M.; Grandér, D.; Morris, K.V. A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* **2013**, *20*, 440–446. [[CrossRef](#)] [[PubMed](#)]
76. Stathopoulou, A.; Chhetri, J.B.; Ambrose, J.C.; Estève, P.-O.; Ji, L.; Erdjument-Bromage, H.; Zhang, G.; Neubert, T.A.; Pradhan, S.; Herrero, J.; et al. A novel requirement for DROSHA in maintenance of mammalian CG methylation. *Nucleic Acids Res.* **2017**, *45*, 9398–9412. [[CrossRef](#)]
77. Zhang, G.; Estève, P.-O.; Chin, H.G.; Terragni, J.; Dai, N.; Corrêa, I.R.; Pradhan, S. Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation. *Nucleic Acids Res.* **2015**, *43*, 6112–6124. [[CrossRef](#)]

78. Thompson, P.J.; Macfarlan, T.S.; Lorincz, M.C. Long terminal repeats: From parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* **2016**, *62*, 766–776. [[CrossRef](#)]
79. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)]
80. Rodriguez-Terrones, D.; Torres-Padilla, M.-E. Nimble and ready to mingle: Transposon outbursts of early development. *Trends Genet. TIG* **2018**, *34*, 806–820. [[CrossRef](#)]
81. Ernst, C.; Odom, D.T.; Kutter, C. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nat. Commun.* **2017**, *8*, 1411. [[CrossRef](#)] [[PubMed](#)]
82. Turelli, P.; Castro-Diaz, N.; Marzetta, F.; Kapopoulou, A.; Raclot, C.; Duc, J.; Tieng, V.; Quenneville, S.; Trono, D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **2014**, *24*, 1260–1270. [[CrossRef](#)] [[PubMed](#)]
83. Kuramochi-Miyagawa, S.; Watanabe, T.; Gotoh, K.; Totoki, Y.; Toyoda, A.; Ikawa, M.; Asada, N.; Kojima, K.; Yamaguchi, Y.; Ijiri, T.W.; et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **2008**, *22*, 908–917. [[CrossRef](#)] [[PubMed](#)]
84. Aravin, A.A.; Sachidanandam, R.; Bourc'his, D.; Schaefer, C.; Pezic, D.; Toth, K.F.; Bestor, T.; Hannon, G.J. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **2008**, *31*, 785–799. [[CrossRef](#)] [[PubMed](#)]
85. Itou, D.; Shiromoto, Y.; Yukiho, S.; Ishii, C.; Nishimura, T.; Ogonuki, N.; Ogura, A.; Hasuwa, H.; Fujihara, Y.; Kuramochi-Miyagawa, S.; et al. Induction of DNA methylation by artificial piRNA production in male germ cells. *Curr. Biol. CB* **2015**, *25*, 901–906. [[CrossRef](#)] [[PubMed](#)]
86. Watanabe, T.; Tomizawa, S.; Mitsuya, K.; Totoki, Y.; Yamamoto, Y.; Kuramochi-Miyagawa, S.; Iida, N.; Hoki, Y.; Murphy, P.J.; Toyoda, A.; et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science* **2011**, *332*, 848–852. [[CrossRef](#)]
87. Liu, J.; Zhang, S.; Cheng, B. Epigenetic roles of PIWI-interacting RNAs (piRNAs) in cancer metastasis (Review). *Oncol. Rep.* **2018**, *40*, 2423–2434. [[CrossRef](#)]
88. Leonhardt, H.; Page, A.W.; Weier, H.U.; Bestor, T.H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **1992**, *71*, 865–873. [[CrossRef](#)]
89. Iida, T.; Suetake, I.; Tajima, S.; Morioka, H.; Ohta, S.; Obuse, C.; Tsurimoto, T. PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes Cells* **2002**, *7*, 997–1007. [[CrossRef](#)]
90. Boehm, E.M.; Washington, M.T. R.I.P. to the PIP: PCNA-binding motif no longer considered specific: PIP motifs and other related sequences are not distinct entities and can bind multiple proteins involved in genome maintenance. *BioEssays* **2016**, *38*, 1117–1122. [[CrossRef](#)]
91. Schermelleh, L.; Haemmer, A.; Spada, F.; Rösing, N.; Meilinger, D.; Rothbauer, U.; Cardoso, M.C.; Leonhardt, H. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* **2007**, *35*, 4301–4312. [[CrossRef](#)] [[PubMed](#)]
92. Sharif, J.; Muto, M.; Takebayashi, S.; Suetake, I.; Iwamatsu, A.; Endo, T.A.; Shinga, J.; Mizutani-Koseki, Y.; Toyoda, T.; Okamura, K.; et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **2007**, *450*, 908–912. [[CrossRef](#)] [[PubMed](#)]
93. Bostick, M.; Kim, J.K.; Estève, P.-O.; Clark, A.; Pradhan, S.; Jacobsen, S.E. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **2007**, *317*, 1760–1764. [[CrossRef](#)] [[PubMed](#)]
94. Achour, M.; Jacq, X.; Rondé, P.; Alhosin, M.; Charlot, C.; Chataigneau, T.; Jeanblanc, M.; Macaluso, M.; Giordano, A.; Hughes, A.D.; et al. The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of *VEGF* gene expression. *Oncogene* **2008**, *27*, 2187–2197. [[CrossRef](#)] [[PubMed](#)]
95. Bashtrykov, P.; Jankevicius, G.; Jurkowska, R.Z.; Ragozin, S.; Jeltsch, A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J. Biol. Chem.* **2014**, *289*, 4106–4115. [[CrossRef](#)] [[PubMed](#)]
96. Berkyurek, A.C.; Suetake, I.; Arita, K.; Takeshita, K.; Nakagawa, A.; Shirakawa, M.; Tajima, S. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J. Biol. Chem.* **2014**, *289*, 379–386. [[CrossRef](#)] [[PubMed](#)]

97. Smets, M.; Link, S.; Wolf, P.; Schneider, K.; Solis, V.; Ryan, J.; Meilinger, D.; Qin, W.; Leonhardt, H. DNMT1 mutations found in HSNIE patients affect interaction with UHRF1 and neuronal differentiation. *Hum. Mol. Genet.* **2017**, *26*, 1522–1534. [[CrossRef](#)] [[PubMed](#)]
98. Ishiyama, S.; Nishiyama, A.; Saeki, Y.; Moritsugu, K.; Morimoto, D.; Yamaguchi, L.; Arai, N.; Matsumura, R.; Kawakami, T.; Mishima, Y.; et al. Structure of the Dnmt1 reader module complexed with a unique two-mono-ubiquitin mark on histone H3 reveals the basis for DNA methylation maintenance. *Mol. Cell* **2017**, *68*, 350–360.e7. [[CrossRef](#)]
99. Xu, C.; Corces, V.G. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science* **2018**, *359*, 1166–1170. [[CrossRef](#)]
100. Charlton, J.; Downing, T.L.; Smith, Z.D.; Gu, H.; Clement, K.; Pop, R.; Akopian, V.; Klages, S.; Santos, D.P.; Tsankov, A.M.; et al. Global delay in nascent strand DNA methylation. *Nat. Struct. Mol. Biol.* **2018**, *25*, 327–332. [[CrossRef](#)]
101. Ferry, L.; Fournier, A.; Tsusaka, T.; Adelmant, G.; Shimazu, T.; Matano, S.; Kirsh, O.; Amouroux, R.; Dohmae, N.; Suzuki, T.; et al. Methylation of DNA ligase 1 by G9a/GLP recruits UHRF1 to replicating DNA and regulates DNA methylation. *Mol. Cell* **2017**, *67*, 550–565.e5. [[CrossRef](#)] [[PubMed](#)]
102. Zhou, T.; Xiong, J.; Wang, M.; Yang, N.; Wong, J.; Zhu, B.; Xu, R.-M. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell* **2014**, *54*, 879–886. [[CrossRef](#)] [[PubMed](#)]
103. Mortusewicz, O.; Schermelleh, L.; Walter, J.; Cardoso, M.C.; Leonhardt, H. Recruitment of DNA methyltransferase I to DNA repair sites. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8905–8909. [[CrossRef](#)] [[PubMed](#)]
104. Ruzov, A.; Shorning, B.; Mortusewicz, O.; Dunican, D.S.; Leonhardt, H.; Meehan, R.R. MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development* **2009**, *136*, 2277–2286. [[CrossRef](#)] [[PubMed](#)]
105. Cuozzo, C.; Porcellini, A.; Angrisano, T.; Morano, A.; Lee, B.; Di Pardo, A.; Messina, S.; Iuliano, R.; Fusco, A.; Santillo, M.R.; et al. DNA damage, homology-directed repair, and DNA methylation. *PLoS Genet.* **2007**, *3*, e110. [[CrossRef](#)]
106. Morano, A.; Angrisano, T.; Russo, G.; Landi, R.; Pezone, A.; Bartollino, S.; Zuchegna, C.; Babbio, F.; Bonapace, I.M.; Allen, B.; et al. Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic Acids Res.* **2014**, *42*, 804–821. [[CrossRef](#)]
107. Ha, K.; Lee, G.E.; Pali, S.S.; Brown, K.D.; Takeda, Y.; Liu, K.; Bhalla, K.N.; Robertson, K.D. Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum. Mol. Genet.* **2011**, *20*, 126–140. [[CrossRef](#)]
108. Tian, Y.; Paramasivam, M.; Ghosal, G.; Chen, D.; Shen, X.; Huang, Y.; Akhter, S.; Legerski, R.; Chen, J.; Seidman, M.M.; et al. UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. *Cell Rep.* **2015**, *10*, 1957–1966. [[CrossRef](#)]
109. Liang, C.-C.; Zhan, B.; Yoshikawa, Y.; Haas, W.; Gygi, S.P.; Cohn, M.A. UHRF1 is a sensor for DNA interstrand crosslinks and recruits FANCD2 to initiate the Fanconi anemia pathway. *Cell Rep.* **2015**, *10*, 1947–1956. [[CrossRef](#)]
110. Miotto, B.; Chibi, M.; Xie, P.; Koundrioukoff, S.; Moolman-Smook, H.; Pugh, D.; Debatisse, M.; He, F.; Zhang, L.; Defossez, P.-A. The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. *Cell Rep.* **2014**, *7*, 575–587. [[CrossRef](#)]
111. Van Houten, B.; Santa-Gonzalez, G.A.; Camargo, M. DNA repair after oxidative stress: Current challenges. *Curr. Opin. Toxicol.* **2018**, *7*, 9–16. [[CrossRef](#)] [[PubMed](#)]
112. O'Hagan, H.M.; Wang, W.; Sen, S.; Destefano Shields, C.; Lee, S.S.; Zhang, Y.W.; Clements, E.G.; Cai, Y.; Van Neste, L.; Easwaran, H.; et al. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell* **2011**, *20*, 606–619. [[CrossRef](#)] [[PubMed](#)]
113. Ding, N.; Bonham, E.M.; Hannon, B.E.; Amick, T.R.; Baylin, S.B.; O'Hagan, H.M. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. *J. Mol. Cell Biol.* **2016**, *8*, 244–254. [[CrossRef](#)] [[PubMed](#)]
114. Maiuri, A.R.; Peng, M.; Podicheti, R.; Sriramkumar, S.; Kamplain, C.M.; Rusch, D.B.; DeStefano Shields, C.E.; Sears, C.L.; O'Hagan, H.M. Mismatch repair proteins initiate epigenetic alterations during inflammation-driven tumorigenesis. *Cancer Res.* **2017**, *77*, 3467–3478. [[CrossRef](#)] [[PubMed](#)]

115. Gu, T.; Lin, X.; Cullen, S.M.; Luo, M.; Jeong, M.; Estecio, M.; Shen, J.; Hardikar, S.; Sun, D.; Su, J.; et al. DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* **2018**, *19*, 88. [[CrossRef](#)] [[PubMed](#)]
116. Noh, K.-M.; Wang, H.; Kim, H.R.; Wenderski, W.; Fang, F.; Li, C.H.; Dewell, S.; Hughes, S.H.; Melnick, A.M.; Patel, D.J.; et al. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol. Cell* **2015**, *59*, 89–103. [[CrossRef](#)] [[PubMed](#)]
117. Holtzman, L.; Gersbach, C.A. Editing the epigenome: Reshaping the genomic landscape. *Annu. Rev. Genomics Hum. Genet.* **2018**, *19*, 43–71. [[CrossRef](#)]
118. Di Ruscio, A.; Ebralidze, A.K.; Benoukraf, T.; Amabile, G.; Goff, L.A.; Terragni, J.; Figueroa, M.E.; De Figueiredo Pontes, L.L.; Alberich-Jorda, M.; Zhang, P.; et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **2013**, *503*, 371–376. [[CrossRef](#)]
119. Pal, S.; Tyler, J.K. Epigenetics and aging. *Sci. Adv.* **2016**, *2*, e1600584. [[CrossRef](#)]

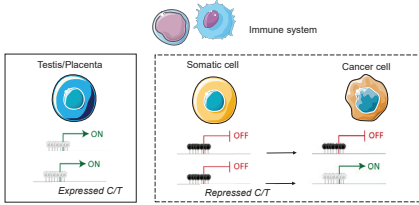


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Marthe Laisné, Mathieu Vanderstraete, Olivier Kirsh, Laure Ferry, Pierre-Antoine Defossez
CNRS UMR7216, University Paris-Diderot, Paris, France

INTRODUCTION:

Cancer/Testis (C/T) genes are epigenetically restricted to immuno-privileged sites, and reactivated in 40% of solid tumors.



Can we use epigenetic drugs to reactivate C/T genes in cancer, and enhance anti-cancer immunotherapy?

Cancer/Testis (C/T) genes are normally epigenetically restricted to germ cells, but can be also reactivated in various types of cancer^(1,2, 3). These genes often encode antigens that are immunogenic in cancer patients, raising the possibility of their usage as biomarkers and as targets for immunotherapies^(4, 5).

Our central issue is the control of Cancer/Testis genes expression. Which epigenetic mechanisms repress their expression in non-transformed cells? Which anomalies lead to their re-expression in tumor cells? Can we manipulate their expression in tumors in order to facilitate their destruction by the immune system?

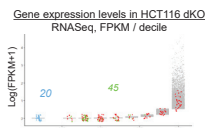
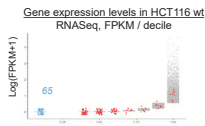
Using colorectal cancers as working model, and more precisely HCT116 cell lines, we have made a chemical screen which allowed the identification of some drugs (different kinase inhibitors) inducing Cancer/Testis genes re-expression. We combine bioinformatics approaches with cellular biology and genomic for understand the molecular and cellular mechanisms monitoring these re-expression, and to clarify the consequences for tumor immune microenvironments.

RESULTS:

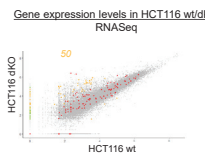
I. Identification of repressed C/T genes in HCT116 model, as candidates for epigenetic reactivation:

A. Which are the repressed C/T genes, inducible by demethylation?

1. Identification of 45 non-expressed C/T genes, methylation sensitive in HCT116 Dnm1^{-/-}; Dnmt3b^{-/-} (dKO):

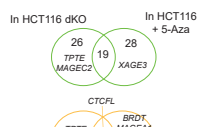


2. Identification of 50 differentially expressed C/T in HCT116 dKO:



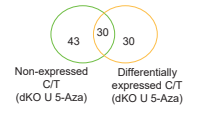
3. Same analyze in HCT116 treated with a demethylation agent (5-Aza):

Comparison between two RNASeq datasets with demethylation treatment:



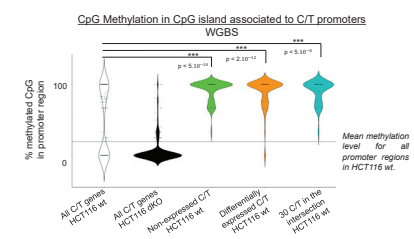
4. 103 C/T genes are predicted to be repressed and sensitive to demethylation in HCT116.

Comparison of the two transcriptomic analyzes (dKO / Aza):

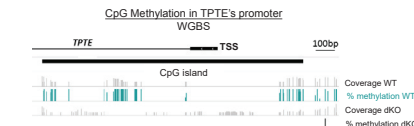


B. Is these genes are actually methylated in their promoters?

1. The two methods allow the selection of repressed C/T genes with highly significant hyper-methylation in CpG island :



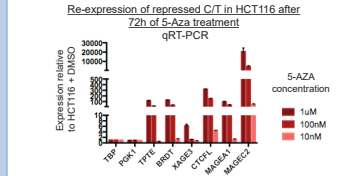
2. We can pick up interesting repressed C/T genes and check for their methylation status:



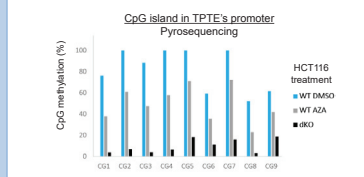
Selection of the most promising candidates from these bioinformatics analyzes for experimental validation.

C. Can we confirm these bioinformatics predictions?

1. All tested C/T genes are expressed as we expected:



2. All tested C/T genes are methylated as we expected:



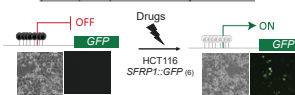
Identification & validation of C/T genes which are:
- Only expressed in testis in the adult ;
- Hyper-methylated in their promoter ;
- Inducible by demethylation in HCT116 colon cancer cell line

II. Identification of potent novel drugs for C/T genes reactivation, by targeting signalic kinase pathways:

A. Are drugs reactivated a methylated Tumor Suppressor Gene, SFRP1?

In a chemical screen (421 kinases inhibitors), identification of drugs which reactivated SFRP1. (Mathieu Vanderstraete)

Model and screening's strategy: a methylated tumor suppressor gene (SFRP1) as reporter of demethylation events



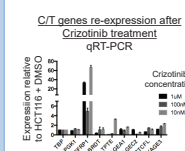
List of hits and their targets:

Drugs	Targets	In vivo tests (mice)	Clinical trials
Alisertib	Aurora	+	Phase 3
AST-1306	EGFR	+	no
AT7519	CDK	+	Phase 2
AT9283	Aurora	+	Phase 2
Barasertib	Aurora	+	Phase 1
BI 2536	PLK	+	Phase 2
Crizotinib	ALK, c-Met	+	FDA approved
GSK461364	PLK	+	Phase 1
JNJ-7706621	Aurora, CDK	+	no
Rigosertib	PLK	+	Phase 3
SB216763	GSK-3	+	no
SNS-032	CDK	+	Phase 1
ZM 447439	Aurora	+	no

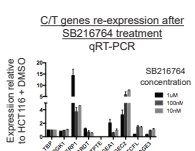
There is a link between signalic kinases and SFRP1 methylation in HCT116 model.
Can we reactivate repressed C/T genes with these compounds?

B. Are these drugs able to reactivate repressed C/T genes?

1. Crizotinib has a specific effect on SFRP1.

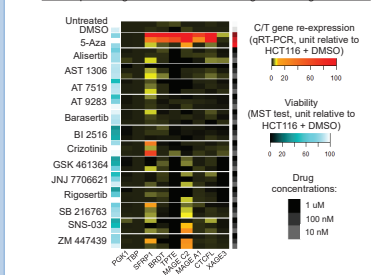


2. SB216764 induces SFRP1 and C/T genes.



3. Preliminary results on 13 drugs from the screen: (n=1, need to be replicated)

Heatmap showing C/T re-activation following 72h of drugs' treatment



FUTURE PERSPECTIVES:

A. Which are the causes of C/T gene re-expression in colorectal cancer cells?

Hypothesis: Signaling pathways are involved in maintenance of C/T gene methylation and transcriptional repression.

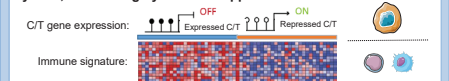
1. Which are the most promising drugs for C/T gene reactivation? 2. Which are the mechanism involved?



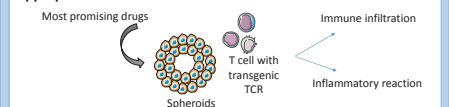
B. Which are the consequences of C/T gene expression on immune microenvironment?

Hypothesis: C/T gene expression could be a marker of abnormality for the immune system, and could initiate an immune response against the tumor.

1. C/T gene expressing-tumor are low infiltrated by the immune system, and / or highly immunosuppressive.



2. C/T gene reactivation in cancer cells are able to activate appropriate immune cells.



Reactivation of epigenetically silenced genes in cancer: Bioinformatics analysis to generate functional hypotheses

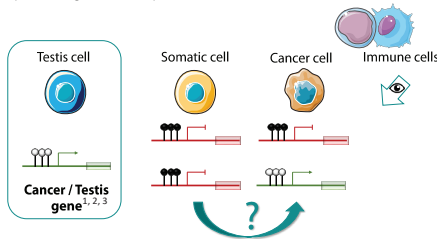
Marthe Laisné (marthe.laisne@gmail.com), Olivier Kirsh, Pierre-Antoine Defossez

University Paris 7 Denis Diderot, Epigenetic and cell fate, UMR 7216 CNRS, 75013 Paris

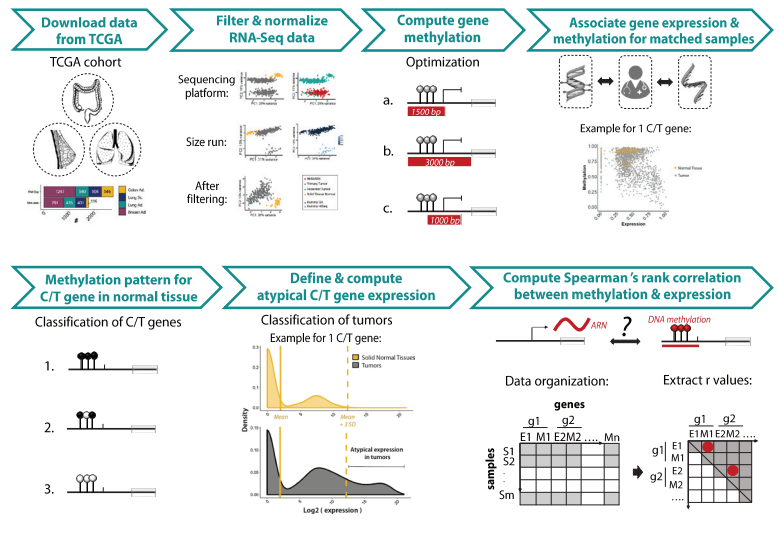
1 Purpose

Epigenomic landscape is remodeled during tumorigenesis. These modifications, like DNA methylation gains and losses, could lead to abnormal activation of epigenetically silenced genes in normal tissues. Our main objective is to generate functional hypotheses on mechanisms involve in the control of DNA methylation pattern. We use The Cancer Genome Atlas (TCGA) data to explore the relationship between DNA methylation and gene expression. In a first instance, we are focused on Cancer/Testis genes as a proof of concept of our approach.

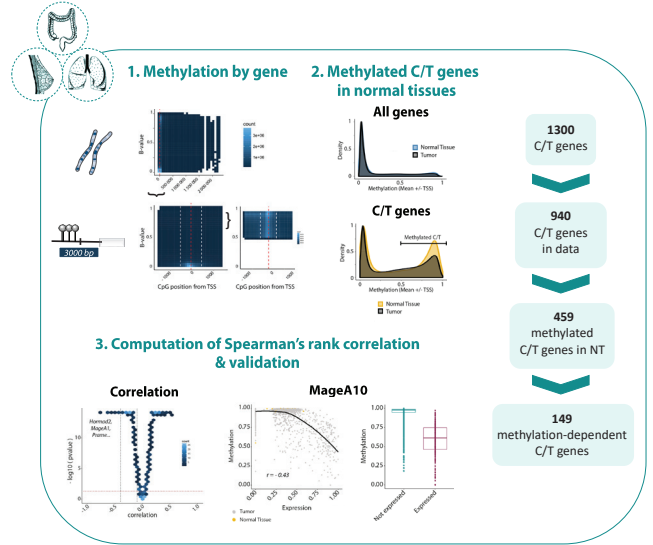
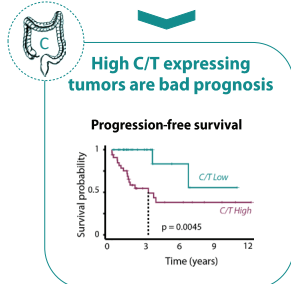
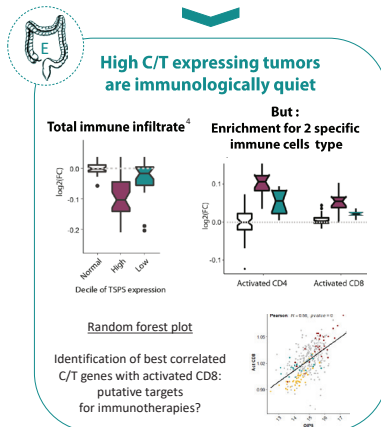
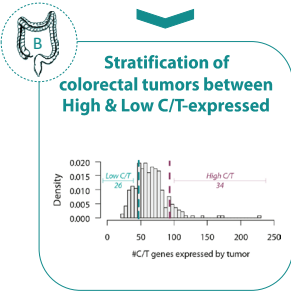
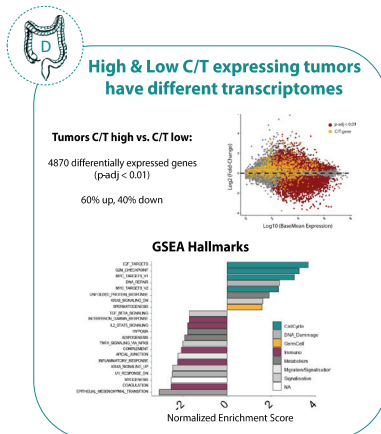
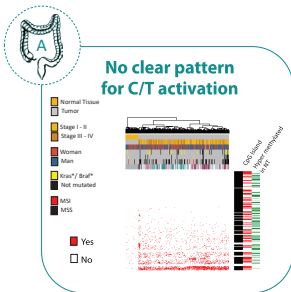
1. We want to better characterize the patterns of C/T genes expression in cancer, and to classify tumors for C/T gene expression. Ultimately, we want to identify potent causal factors of C/T gene expression in tumors, and test them functionally.
2. More generally, we want to define new gene-set of DNA-methylation dependent genes, re-expressed in cancer.



2 Methods



3 Results



4 Conclusion & Perspectives

1. At least in the colorectal TCGA cohort, C/T gene expression is correlated with aggressiveness of tumors, maybe linked to a stem-cell like state, and with a depleted microenvironment in immune cells. These preliminary results encourage us to go further into the mechanisms underlying C/T genes activation.
2. We have developed a method to select genes which are highly methylation-dependent in their regulation. Next, we would like to study further the genomic and transcriptomic context which support these epigenomic alterations.
3. We are working on developing and looking for methods to identify genes whose expressions in cancer is linked to epigenomic alteration.

1. Wang, Nat. Communication 2016 ; 2. Almeida, NMS 2009 ; 3. Rousseau, Sci. Transl. Med. 2013 ; 4. Angelova, Genome Biology 2015



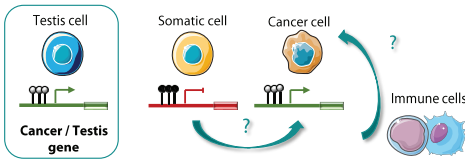
Activation of epigenetically silenced genes in cancer: Bioinformatics analysis to generate functional hypotheses

Introduction

Epigenetic landscape and transcription networks are remodeled during tumorigenesis. These modifications, like DNA methylation gains and losses, can lead to abnormal activation of gene silenced in normal cells. Our main objective is to generate functional hypotheses on mechanisms involved in the control of DNA methylation, in the context of tumorigenesis.

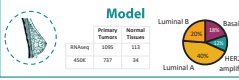
We are using The Cancer Genome Atlas (TCGA) data to develop models of aberrant activation of genes, and to find correlations between these abnormal activations and epigenetic and/or transcriptional pathways activities. We are mainly focusing on Cancer/Testis (C/T) genes, as they are frequently activated in cancers.

1. We want to better characterize patterns of aberrant C/T gene activation in cancers, and to understand the consequences of C/T gene activation for tumor development.
2. Moreover, we want to identify potent causal factors of C/T gene expression in tumors, and test them functionally.



1 A bioinformatic screen identifies 158 C/T genes abnormally activated in breast cancers

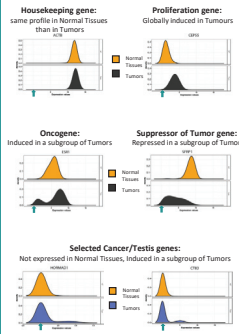
Three different studies define genes with a Testis-specific expression pattern



Definition of rules to identify abnormally expressed genes in tumors

1. Unimodal profile in Normal Breast Tissues: derivative sign of Normal Tissue density changes only once.
2. Multimodal profile in Breast Tumors: derivative sign of Tumor density changes two or more times.

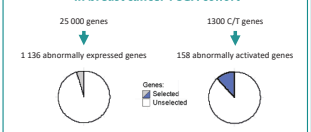
Examples of gene expression profiles



Development & optimization of an automated pipeline to identify aberrantly expressed genes

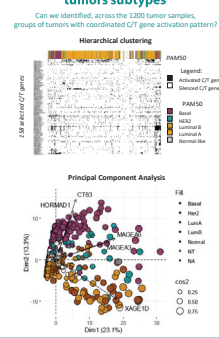
- For each gene:
1. Calculate the **kernel density estimation** of the gene expression in Normal Tissues ; and in Tumors
 2. Calculate the **derivative** of each densities
 3. Sum the number of sign changes for each derivatives, i.e. the **number of variations** in the density curve
- ⇒ Computational time: less than 2min for 2200 genes in 12000 samples
⇒ Choose fixed parameters: smoothing bandwidth to be used in Kernel density estimation

158 Cancer/Testis genes are abnormally expressed in breast cancer TCGA cohort

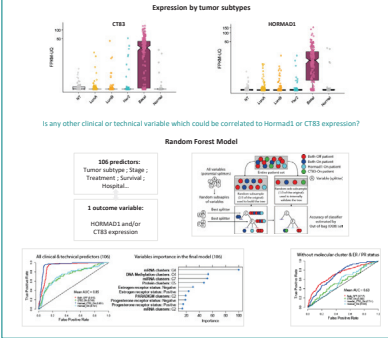


2 Two C/T genes, *HORMAD1* & *CT83*, are markers of basal breast cancer cells

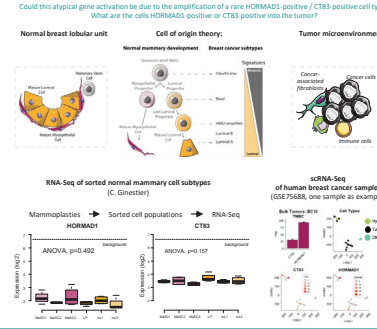
These 158 C/T segregate with tumors subtypes



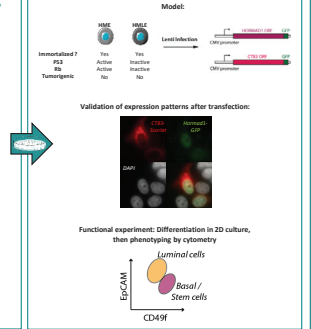
HORMAD1 and *CT83* are specifically induced in the basal subtype



HORMAD1 and *CT83* are activated in cancer cells, after transformation

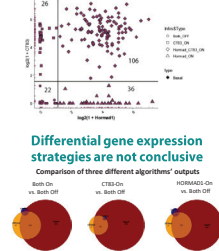


Are *HORMAD1* and *CT83* involved in basal breast tumor oncogenesis?

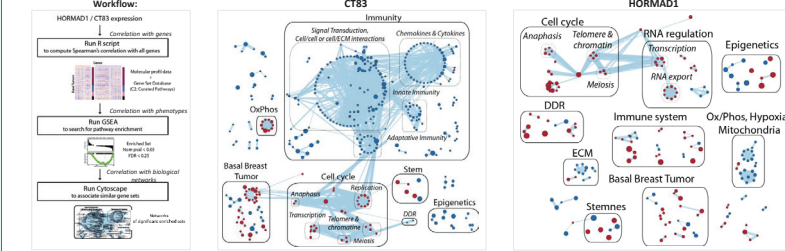


3 What are the consequences of *HORMAD1* & *CT83* activation?

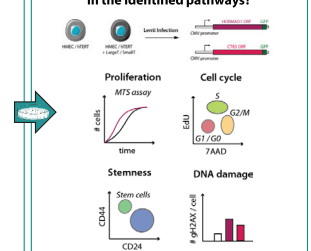
Model: basal breast tumors



Interaction maps of biological functions correlated with *HORMAD1* or *CT83* expression

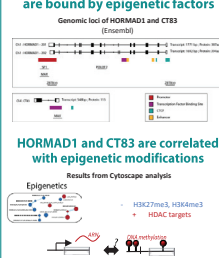


Are *HORMAD1* and *CT83* involved in the identified pathways?

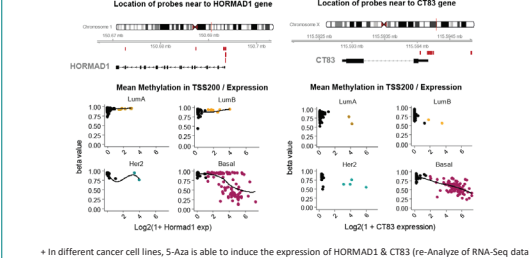


4 What are the mechanisms of *HORMAD1* & *CT83* activation?

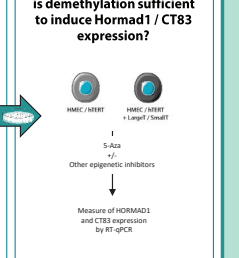
HORMAD1 and *CT83* promoters are bound by epigenetic factors



In cancers, demethylation is sufficient to induce *HORMAD1* & *CT83* expression



In non-cancerous cells, is demethylation sufficient to induce *HORMAD1* / *CT83* expression?



Conclusion

- We have developed a method to automatically detect atypical gene expression in tumors.
- 158 Cancer/Testis genes have been identified as abnormally activated in breast tumors, compared to normal breast tissues.
- Two Cancer/Testis genes, *HORMAD1* and *CT83*, are specific markers of basal breast tumor cells.
- Hypotheses about their functions and their regulatory mechanisms will be further tested, using cellular models of immortalized cell lines.

BIBLIOGRAPHIE



- Achinger-Kawecka, J., Valdes-Mora, F., Luu, P.-L., Giles, K.A., Caldon, C.E., Qu, W., Nair, S., Soto, S., Locke, W.J., Yeo-Teh, N.S., et al. (2020). Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. *Nature Communications* 11, 320.
- Adélaïde, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D., et al. (2007). Integrated Profiling of Basal and Luminal Breast Cancers. *Cancer Res* 67, 11565–11575.
- Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213.
- Akashi, M., Yamaguchi, R., Kusano, H., Obara, H., Yamaguchi, M., Toh, U., Akiba, J., Kakuma, T., Tanaka, M., Akagi, Y., et al. (2020). Diverse histomorphology of HER2-positive breast carcinomas based on differential ER expression. *Histopathology* 76, 560–571.
- Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. *Nature Reviews Molecular Cell Biology* 13, 153–167.
- Alberti, L., Renaud, S., Losi, L., Leyvraz, S., and Benhattar, J. (2014). High expression of hTERT and stemness genes in BORIS/CTCF positive cells isolated from embryonic cancer cells. *PLoS One* 9, e109921.
- Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat Med* 21, 846–853.
- Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., et al. (2004). Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 6, 17–32.
- Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* 19, 229–244.
- van Amerongen, R., Bowman, A.N., and Nusse, R. (2012). Developmental stage and time dictate the fate of Wnt/ β -catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* 11, 387–400.
- Attar, N., Campos, O.A., Vogelauer, M., Cheng, C., Xue, Y., Schmollinger, S., Salwinski, L., Mallipeddi, N.V., Boone, B.A., Yen, L., et al. (2020). The histone H3-H4 tetramer is a copper reductase enzyme. *Science* 369, 59–64.
- Audia, J.E., and Campbell, R.M. (2016). Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol* 8.
- Aung, P.P., Oue, N., Mitani, Y., Nakayama, H., Yoshida, K., Noguchi, T., Bosserhoff, A.K., and Yasui, W. (2006). Systematic search for gastric cancer-specific genes based on SAGE data: melanoma inhibitory activity and matrix metalloproteinase-10 are novel prognostic factors in patients with gastric cancer. *Oncogene* 25, 2546–2557.
- Bachman, K.E., Argani, P., Samuels, Y., Silliman, N., Ptak, J., Szabo, S., Konishi, H., Karakas, B., Blair, B.G., Lin, C., et al. (2004). The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer Biol Ther* 3, 772–775.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research* 21, 381–395.
- Barter, M.J., Pybus, L., Litherland, G.J., Rowan, A.D., Clark, I.M., Edwards, D.R., Cawston, T.E., and Young, D.A. (2010). HDAC-mediated control of ERK- and PI3K-dependent TGF- β -induced extracellular matrix-regulating genes. *Matrix Biol* 29, 602–612.
- Baylin, S.B., and Jones, P.A. (2016). Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol* 8.
- Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *PNAS* 96, 3969–3974.
- Bhattacharyya, T., Walker, M., Powers, N.R., Brunton, C., Fine, A.D., Petkov, P.M., and Handel, M.A. (2019). Prdm9 and Meiotic Cohesin Proteins Cooperatively Promote DNA Double-Strand Break Formation in

Mammalian Spermatocytes. *Current Biology* 29, 1002-1018.e7.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S., and Di Croce, L. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. *Trends in Genetics* 36, 118–131.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42, 1093–1100.

Bowman, G.D., and Poirier, M.G. (2015). Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem. Rev.* 115, 2274–2295.

Britschgi, A., Duss, S., Kim, S., Couto, J.P., Brinkhaus, H., Koren, S., De Silva, D., Mertz, K.D., Kaup, D., Varga, Z., et al. (2017). Hippo kinases LATS1/2 control human breast cell fate via crosstalk with ER α . *Nature* 541, 541–545.

Briu, L.-M., Maric, C., and Cadoret, J.-C. (2021). Replication Stress, Genomic Instability, and Replication Timing: A Complex Relationship. *International Journal of Molecular Sciences* 22, 4764.

Brooks, M.D., Burness, M.L., and Wicha, M.S. (2015). Therapeutic Implications of Cellular Heterogeneity and Plasticity in Breast Cancer. *Cell Stem Cell* 17, 260–271.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. *Nature Reviews Molecular Cell Biology* 18, 299–314.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.

Caravaca, J.M., Donahue, G., Becker, J.S., He, X., Vinson, C., and Zaret, K.S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev* 27, 251–260.

Carballo, J.A., Johnson, A.L., Sedgwick, S.G., and Cha, R.S. (2008). Phosphorylation of the axial element protein Hop1 by Mec1/Tel1 ensures meiotic interhomolog recombination. *Cell* 132, 758–770.

Casey, A.E., Sinha, A., Singhanian, R., Livingstone, J., Waterhouse, P., Tharmapalan, P., Cruickshank, J., Shehata, M., Drysdale, E., Fang, H., et al. (2018a). Mammary molecular portraits reveal lineage-specific features and progenitor cell vulnerabilities. *J Cell Biol* 217, 2951–2974.

Casey, A.E., Sinha, A., Singhanian, R., Livingstone, J., Waterhouse, P., Tharmapalan, P., Cruickshank, J., Shehata, M., Drysdale, E., Fang, H., et al. (2018b). Mammary molecular portraits reveal lineage-specific features and progenitor cell vulnerabilities. *J Cell Biol* 217, 2951–2974.

Chen, B., Tang, H., Chen, X., Zhang, G., Wang, Y., Xie, X., and Liao, N. (2018). Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag Res* 11, 179–190.

Chen, B., Tang, H., Chen, X., Zhang, G., Wang, Y., Xie, X., and Liao, N. (2019a). Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag Res* 11, 179–190.

Chen, H., Fujioka, M., Jaynes, J.B., and Gregor, T. (2017). Direct visualization of transcriptional activation by physical enhancer–promoter proximity. *BioRxiv* 099523.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature Genetics* 39, 391–396.

Chen, Y.-T., Ross, D.S., Chiu, R., Zhou, X.K., Chen, Y.-Y., Lee, P., Hoda, S.A., Simpson, A.J., Old, L.J., Caballero, O., et al. (2011). Multiple Cancer/Testis Antigens Are Preferentially Expressed in Hormone-Receptor Negative and High-Grade Breast Cancers. *PLOS ONE* 6, e17876.

Chen, Z., Zuo, X., Pu, L., Zhang, Y., Han, G., Zhang, L., Wu, Z., You, W., Qin, J., Dai, X., et al. (2019b). Hypomethylation-mediated activation of cancer/testis antigen KK-LC-1 facilitates hepatocellular carcinoma progression through activating the Notch1/Hes1 signalling. *Cell Prolif.* 52, e12581.

Chew, G.-L., Campbell, A.E., De Neef, E., Sutliff, N.A., Shadle, S.C., Tapscott, S.J., and Bradley, R.K. (2019). DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint

Blockade. *Developmental Cell* 50, 658–671.e7.

Chiang, H.-C., Zhang, X., Li, J., Zhao, X., Chen, J., Wang, H.T.-H., Jatoi, I., Brenner, A., Hu, Y., and Li, R. (2019). BRCA1-associated R-loop affects transcription and differentiation in breast luminal epithelial cells. *Nucleic Acids Res* 47, 5086–5099.

Ciriello, G., Sinha, R., Hoadley, K.A., Jacobsen, A.S., Reva, B., Perou, C.M., Sander, C., and Schultz, N. (2013). The molecular diversity of Luminal A breast tumors. *Breast Cancer Res Treat* 141, 409–420.

Ciriello, G., Gatz, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., et al. (2015). Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 163, 506–519.

Ciró, M., Prosperini, E., Quarto, M., Grazini, U., Walfridsson, J., McBlane, F., Nucifero, P., Pacchiana, G., Capra, M., Christensen, J., et al. (2009). ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. *Cancer Res* 69, 8491–8498.

Clark, S.J., Lee, H.J., Smallwood, S.A., Kelsey, G., and Reik, W. (2016). Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* 17. Colot, V., and Rossignol, J.L. (1999). Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21, 402–411.

Copeland, R.A., Solomon, M.E., and Richon, V.M. (2009). Protein methyltransferases as a target class for drug discovery. *Nat Rev Drug Discov* 8, 724–732.

Costa, E., Ferreira-Gonçalves, T., Chasqueira, G., Cabrita, A.S., Figueiredo, I.V., and Reis, C.P. (2020). Experimental Models as Refined Translational Tools for Breast Cancer Research. *Scientia Pharmaceutica* 88, 32.

Curigliano, G., Viale, G., Ghioni, M., Jungbluth, A.A., Bagnardi, V., Spagnoli, G.C., Neville, A.M., Nolè, F., Rotmensz, N., and Goldhirsch, A. (2011). Cancer–testis antigen expression in triple-negative breast cancer. *Annals of Oncology* 22, 98–103.

Dabin, J., Fortuny, A., and Polo, S.E. (2016). Epigenome maintenance in response to DNA damage. *Mol Cell* 62, 712–727.

Daniel, K., Lange, J., Hached, K., Fu, J., Anastassiadis, K., Roig, I., Cooke, H.J., Stewart, A.F., Wassmann, K., Jasin, M., et al. (2011). Meiotic homologous chromosome alignment and its surveillance are controlled by mouse *HORMAD1*. *Nat Cell Biol* 13, 599–610.

Dawson, M.A. (2017a). The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* 355, 1147–1152.

Dawson, M.A., and Kouzarides, T. (2012). Cancer Epigenetics: From Mechanism to Therapy. *Cell* 150, 12–27.

Deblois, G., Tonekaboni, S.A.M., Grillo, G., Martinez, C., Kao, Y.I., Tai, F., Ettayebi, I., Fortier, A.-M., Savage, P., Fedor, A.N., et al. (2020). Epigenetic Switch–Induced Viral Mimicry Evasion in Chemotherapy-Resistant Breast Cancer. *Cancer Discov* 10, 1312–1329.

Dedeurwaerder, S., Desmedt, C., Calonne, E., Singhal, S.K., Haibe-Kains, B., Defrance, M., Michiels, S., Volkmar, M., Deplus, R., Luciani, J., et al. (2011). DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* 3, 726–741.

Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology* 20, 69–84.

Dravis, C., Chung, C.-Y., Lytle, N.K., Herrera-Valdez, J., Luna, G., Trejo, C.L., Reya, T., and Wahl, G.M. (2018a). Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell* 34, 466–482.e6.

Dravis, C., Chung, C.-Y., Lytle, N.K., Herrera-Valdez, J., Luna, G., Trejo, C.L., Reya, T., and Wahl, G.M. (2018b). Epigenetic and transcriptomic profiling of mammary gland development and tumor models disclose regulators of cell state plasticity. *Cancer Cell* 34, 466–482.e6.

Du, Q., Bert, S.A., Armstrong, N.J., Caldon, C.E., Song, J.Z., Nair, S.S., Gould, C.M., Luu, P.-L., Peters, T.,

- Khoury, A., et al. (2019). Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature Communications* 10, 416.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Dunn, J., and Rao, S. (2017). Epigenetics and immunotherapy: The current state of play. *Mol Immunol* 87, 227–239.
- Easwaran, H., Tsai, H.-C., and Baylin, S.B. (2014). Cancer epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Mol Cell* 54, 716–727.
- Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal Instability and Tumors Promoted by DNA Hypomethylation. *Science* 300, 455–455.
- Elster, N., Collins, D.M., Toomey, S., Crown, J., Eustace, A.J., and Hennessy, B.T. (2015). HER2-family signalling mechanisms, clinical implications and targeting in breast cancer. *Breast Cancer Res Treat* 149, 5–15.
- Emens, L.A. (2018). Breast Cancer Immunotherapy: Facts and Hopes. *Clin Cancer Res* 24, 511–520.
- Epping, M.T., Wang, L., Edel, M.J., Carlée, L., Hernandez, M., and Bernards, R. (2005). The Human Tumor Antigen PRAME Is a Dominant Repressor of Retinoic Acid Receptor Signaling. *Cell* 122, 835–847.
- Escobar, T.M., Loyola, A., and Reinberg, D. (2021). Parental nucleosome segregation and the inheritance of cellular identity. *Nature Reviews Genetics* 1–14.
- Esteller, M. (2004). Aberrant dna methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.* 45, 629–656.
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics* 8, 286–298.
- Esteller, M. (2008). *Epigenetics in Biology and Medicine* (CRC Press).
- Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., et al. (2000). Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors. *JNCI: Journal of the National Cancer Institute* 92, 564–569.
- Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., Macgrogan, G., Bergh, J., Cameron, D., Goldstein, D., et al. (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 24, 4660–4671.
- Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. *Nature Reviews Cancer* 4, 143–153.
- Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks of cancer. *Science* 357.
- Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 37, 391–400.
- Fu, N.Y., Rios, A.C., Pal, B., Soetanto, R., Lun, A.T.L., Liu, K., Beck, T., Best, S.A., Vaillant, F., Bouillet, P., et al. (2015). EGF-mediated induction of Mcl-1 at the switch to lactation is essential for alveolar cell survival. *Nature Cell Biology* 17, 365–375.
- Fukuyama, T., Hanagiri, T., Takenoyama, M., Ichiki, Y., Mizukami, M., So, T., Sugaya, M., So, T., Sugio, K., and Yasumoto, K. (2006). Identification of a new cancer/germline gene, KK-LC-1, encoding an antigen recognized by autologous CTL induced on human lung adenocarcinoma. *Cancer Res.* 66, 4922–4928.
- Fukuyama, T., Futawatari, N., Yamamura, R., Yamazaki, T., Ichiki, Y., Ema, A., Ushiku, H., Nishi, Y., Takahashi, Y., Otsuka, T., et al. (2018). Expression of KK-LC-1, a cancer/testis antigen, at non-tumour sites of the stomach carrying a tumour. *Sci Rep* 8.
- Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L.-K., Srinivasan, M., Zhang, H., Meric-Bernstam, F., and Navin, N. (2017). Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast

cancer. *Nat Commun* 8, 228.

Gao, Y., Kardos, J., Yang, Y., Tamir, T.Y., Mutter-Rottmayer, E., Weissman, B., Major, M.B., Kim, W.Y., and Vaziri, C. (2018). The Cancer/Testes (CT) Antigen HORMAD1 promotes Homologous Recombinational DNA Repair and Radioresistance in Lung adenocarcinoma cells. *Sci Rep* 8, 15304.

Garcia-Martinez, L., Zhang, Y., Nakata, Y., Chan, H.L., and Morey, L. (2021). Epigenetic mechanisms in breast cancer therapy and resistance. *Nature Communications* 12, 1786.

Garrido-Castro, A.C., Lin, N.U., and Polyak, K. (2019). Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov* 9, 176–198.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* 12, 36–47.

Gatza, M.L., Silva, G.O., Parker, J.S., Fan, C., and Perou, C.M. (2014). An integrated genomics approach identifies drivers of proliferation in luminal subtype human breast cancer. *Nat Genet* 46, 1051–1059.

Gerratana, L., Basile, D., Buono, G., Placido, S., Giuliano, M., Minichillo, S., Coinu, A., Martorana, F., Santo, I., Mastro, L., et al. (2018). Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treatment Reviews* 68.

Gibbs, Z.A., and Whitehurst, A.W. (2018). Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors. *Trends Cancer* 4, 701–712.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell* 153, 1149–1163.

Godoy-Ortiz, A., Sanchez-Muñoz, A., Chica Parrado, M.R., Álvarez, M., Ribelles, N., Rueda Dominguez, A., and Alba, E. (2019). Deciphering HER2 Breast Cancer Disease: Biological and Clinical Implications. *Front. Oncol.* 0.

Gökbuget, D., and Blelloch, R. (2019). Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development* 146.

Graff, J.R., Gabrielson, E., Fujii, H., Baylin, S.B., and Herman, J.G. (2000). Methylation Patterns of the E-cadherin 5' CpG Island Are Unstable and Reflect the Dynamic, Heterogeneous Loss of E-cadherin Expression during Metastatic Progression*. *Journal of Biological Chemistry* 275, 2727–2732.

Granit, R.Z., Masury, H., Condiotti, R., Fixler, Y., Gabai, Y., Glikman, T., Dalin, S., Winter, E., Nevo, Y., Carmon, E., et al. (2018). Regulation of Cellular Heterogeneity and Rates of Symmetric and Asymmetric Divisions in Triple-Negative Breast Cancer. *Cell Reports* 24, 3237–3250.

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, 590–607.

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Rey, F., Frenoy, O., et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nature Genetics* 51, 1060–1066.

Gruver, A.M., Portier, B.P., and Tubbs, R.R. (2011). Molecular pathology of breast cancer: the journey from traditional practice toward embracing the complexity of a molecular classification. *Arch Pathol Lab Med* 135, 544–557.

Gu, Y., Wang, C., Zhu, R., Yang, J., Yuan, W., Zhu, Y., Zhou, Y., Qin, N., Shen, H., Ma, H., et al. (2021). The cancer-testis gene, MEIOB, sensitizes triple-negative breast cancer to PARP1 inhibitors by inducing homologous recombination deficiency. *Cancer Biol Med* 18, 74–87.

Guarente, L., and Picard, F. (2005). Calorie restriction--the SIR2 connection. *Cell* 120, 473–482. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.

- Hauer, M.H., and Gasser, S.M. (2017). Chromatin and nucleosome dynamics in DNA damage and repair. *Genes Dev.* 31, 2204–2221.
- Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491.
- Hekselman, I., and Yeger-Lotem, E. (2020). Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nature Reviews Genetics* 21, 137–150.
- Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? *Trends Genet* 27, 389–396.
- Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., et al. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8, R76.
- Hinohara, K., and Polyak, K. (2019). Intratumoral heterogeneity: more than just mutations. *Trends in Cell Biology* 29, 569.
- Hofmann, O., Caballero, O.L., Stevenson, B.J., Chen, Y.-T., Cohen, T., Chua, R., Maher, C.A., Panji, S., Schaefer, U., Kruger, A., et al. (2008). Genome-wide analysis of cancer/testis gene expression. *Proc Natl Acad Sci U S A* 105, 20422–20427.
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
- Holliday, H., Baker, L.A., Junankar, S.R., Clark, S.J., and Swarbrick, A. (2018). Epigenomics of mammary gland development. *Breast Cancer Research* 20, 100.
- Holm, K., Staaf, J., Lauss, M., Aine, M., Lindgren, D., Bendahl, P.-O., Vallon-Christersson, J., Barkardottir, R.B., Höglund, M., Borg, Å., et al. (2016). An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. *Breast Cancer Res.* 18, 27.
- Hong, S.P., Chan, T.E., Lombardo, Y., Corleone, G., Rotmensch, N., Bravaccini, S., Rocca, A., Pruneri, G., McEwen, K.R., Coombes, R.C., et al. (2019). Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. *Nat Commun* 10.
- Hosoya, N., Okajima, M., Kinomura, A., Fujii, Y., Hiyama, T., Sun, J., Tashiro, S., and Miyagawa, K. (2012). Synaptonemal complex protein SYCP3 impairs mitotic recombination by interfering with BRCA2. *EMBO Rep* 13, 44–51.
- Howard, B., and Ashworth, A. (2006). Signalling Pathways Implicated in Early Mammary Gland Morphogenesis and Breast Cancer. *PLoS Genet* 2.
- Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7, 96.
- Iranzo, J., Martincorena, I., and Koonin, E. (2018). Cancer-mutation network and the number and specificity of driver mutations. *Proceedings of the National Academy of Sciences* 115, 201803155.
- Issa, J.-P. (2000). CpG-Island Methylation in Aging and Cancer. In *DNA Methylation and Cancer*, P.A. Jones, and P.K. Vogt, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 101–118.
- Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. *Genes Dev* 28, 2679–2692.
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* 62, 79–91.
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* 33, 245–254.
- Jakovcevski, M., and Akbarian, S. (2012). Epigenetic mechanisms in neurodevelopmental and neurodegenerative disease. *Nat Med* 18, 1194–1204.

Janin, M., and Esteller, M. (2020). Epigenetic Awakening of Viral Mimicry in Cancer. *Cancer Discov* 10, 1258–1260.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.e19.

Jin, W., Chen, L., Chen, Y., Xu, S.-G., Di, G.-H., Yin, W.-J., Wu, J., and Shao, Z.-M. (2010). UHRF1 is associated with epigenetic silencing of BRCA1 in sporadic breast cancer. *Breast Cancer Res Treat* 123, 359–373

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics* 3, 415–428.

Jones, P.A., Issa, J.-P.J., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. *Nature Reviews Genetics* 17, 630–641.

Jovanovic, J., Rønneberg, J.A., Tost, J., and Kristensen, V. (2010). The epigenetics of breast cancer. *Mol Oncol* 4, 242–254.

Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. *Nature Reviews Cancer* 12, 381–385.

Karaayvaz, M., Cristea, S., Gillespie, S.M., Patel, A.P., Mylvaganam, R., Luo, C.C., Specht, M.C., Bernstein, B.E., Michor, F., and Ellisen, L.W. (2018). Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat Commun* 9, 3588.

Karlin, K.L., Mondal, G., Hartman, J.K., Tyagi, S., Kurley, S.J., Bland, C.S., Hsu, T.Y.T., Renwick, A., Fang, J.E., Migliaccio, I., et al. (2014). The oncogenic STP axis promotes triple-negative breast cancer via degradation of the REST tumor suppressor. *Cell Rep* 9, 1318–1332.

Karpf, A.R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res* 65, 8635–8639.

Karthikeyan, S., Waters, I.G., Dennison, L., Chu, D., Donaldson, J., Shin, D.H., Rosen, D.M., Gonzalez-Ericsson, P.I., Sanchez, V., Sanders, M.E., et al. (2021). Hierarchical tumor heterogeneity mediated by cell contact between distinct genetic subclones. *J Clin Invest* 131.

Kaufmann, J., Wentzensen, N., Brinker, T.J., and Grabe, N. (2019). Large-scale in-silico identification of a tumor-specific antigen pool for targeted immunotherapy in triple-negative breast cancer. *Oncotarget* 10, 2515–2529.

Kikuchi, R., Yagi, S., Kusuhara, H., Imai, S., Sugiyama, Y., and Shiota, K. (2010). Genome-wide analysis of epigenetic signatures for kidney-specific transporters. *Kidney Int* 78, 569–577.

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single Cell Sequencing. *Cell* 173, 879–893.e13.

Kleer, C.G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S.A., Ghosh, D., Sewalt, R.G.A.B., Otte, A.P., Hayes, D.F., et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 100, 11606–11611.

Kondo, Y., Fukuyama, T., Yamamura, R., Futawatari, N., Ichiki, Y., Tanaka, Y., Nishi, Y., Takahashi, Y., Yamazaki, H., Kobayashi, N., et al. (2018). Detection of KK-LC-1 Protein, a Cancer/Testis Antigen, in Patients with Breast Cancer. *Anticancer Res.* 38, 5923–5928.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* 128, 693–705.
Kröger, C., Afeyan, A., Mraz, J., Eaton, E.N., Reinhardt, F., Khodor, Y.L., Thiru, P., Bierie, B., Ye, X., Burge, C.B., et al. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci U S A* 116, 7353–7362.

Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* 18, 205–214.

Kumar, R., Ghyselinck, N., Ishiguro, K., Watanabe, Y., Kouznetsova, A., Höög, C., Strong, E., Schimenti, J., Daniel, K., Toth, A., et al. (2015). MEI4 – a central player in the regulation of meiotic DNA double-strand

break formation in the mouse. *J. Cell. Sci.* 128, 1800–1811.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.

Kwei, K.A., Kung, Y., Salari, K., Holcomb, I.N., and Pollack, J.R. (2010). Genomic instability in breast cancer: Pathogenesis and clinical implications. *Mol Oncol* 4, 255–266.

LaFave, L.M., Kartha, V.K., Ma, S., Meli, K., Del Priore, I., Lareau, C., Naranjo, S., Westcott, P.M.K., Duarte, F.M., Sankar, V., et al. (2020). Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* 38, 212-228.e13.

Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., and Defossez, P.-A. (2018). Mechanisms of DNA Methyltransferase Recruitment in Mammals. *Genes* 9, 617.

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 121, 2750–2767.

Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., and Pietenpol, J.A. (2016). Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS One* 11.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* 3, 662–673.

Lindeman, G.J., and Visvader, J.E. (2011). Cell fate takes a slug in BRCA1-associated breast cancer. *Breast Cancer Res* 13, 306.

Lips, E.H., Mulder, L., Oonk, A., van der Kolk, L.E., Hogervorst, F.B.L., Imholz, A.L.T., Wesseling, J., Rodenhuis, S., and Nederlof, P.M. (2013). Triple-negative breast cancer: BRCAness and concordance of clinical features with BRCA1-mutation carriers. *Br J Cancer* 108, 2172–2177.

Liu, K., Newbury, P.A., Glicksberg, B.S., Zeng, W.Z.D., Paithankar, S., Andrechek, E.R., and Chen, B. (2019). Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nat Commun* 10, 2138.

Liu, K., Wang, Y., Zhu, Q., Li, P., Chen, J., Tang, Z., Shen, Y., Cheng, X., Lu, L.-Y., and Liu, Y. (2020). Aberrantly expressed *HORMAD1* disrupts nuclear localization of MCM8–MCM9 complex and compromises DNA mismatch repair in cancer cells. *Cell Death & Disease* 11, 1–15.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C.G., Merajver, S.D., Dontu, G., and Wicha, M.S. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. *Proc Natl Acad Sci U S A* 105, 1680–1685.

Liu, Y., DeBoer, K., de Kretser, D.M., O'Donnell, L., O'Connor, A.E., Merriner, D.J., Okuda, H., Whittle, B., Jans, D.A., Efthymiadis, A., et al. (2015). *LRGUK-1* Is Required for Basal Body and Manchette Function during Spermatogenesis and Male Fertility. *PLoS Genet* 11, e1005090.

Ljungman, M., and Hanawalt, P.C. (1992). Efficient protection against oxidative DNA damage in chromatin. *Mol Carcinog* 5, 264–269.

Locke, W.J., and Clark, S.J. (2012). Epigenome remodelling in breast cancer: insights from an early in vitro model of carcinogenesis. *Breast Cancer Res* 14, 215.

Lord, C.J., and Ashworth, A. (2016). BRCAness revisited. *Nature Reviews Cancer* 16, 110–120.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* 183, 1103-1116.e20.

Mader, C. (2007). The Biology of Cancer. *Yale J Biol Med* 80, 91.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends in Genetics* 27, 465–474.

- Magnani, L., Stoeck, A., Zhang, X., Lánczky, A., Mirabella, A.C., Wang, T.-L., Gyorffy, B., and Lupien, M. (2013). Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *PNAS* 110, E1490–E1499.
- Mahmoud, A.M. (2018). Cancer testis antigens as immunogenic and oncogenic targets in breast cancer. *Immunotherapy* 10, 769–778.
- Maine, E.A., Westcott, J.M., Prechtel, A.M., Dang, T.T., Whitehurst, A.W., and Pearson, G.W. (2016). The cancer-testis antigens SPANX-A/C/D and CTAG2 promote breast cancer invasion. *Oncotarget* 7, 14708–14726.
- Marchiò, C., Annaratone, L., Marques, A., Casorzo, L., Berrino, E., and Sapino, A. (2021). Evolving concepts in HER2 evaluation in breast cancer: Heterogeneity, HER2-low carcinomas and beyond. *Seminars in Cancer Biology* 72, 123–135.
- Marcinkowski, B., Stevanović, S., Helman, S.R., Norberg, S.M., Serna, C., Jin, B., Gkitsas, N., Kadakia, T., Warner, A., Davis, J.L., et al. (2019). Cancer targeting by TCR gene-engineered T cells directed against Kita-Kyushu Lung Cancer Antigen-1. *J Immunother Cancer* 7.
- Marine, J.-C., Dawson, S.-J., and Dawson, M.A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. *Nature Reviews Cancer* 20, 743–756.
- Marshall, E. (2014). Dare to Do Less. *Science* 343, 1454–1456.
- Marsolier, J., Prompsy, P., Durand, A., Lyne, A.-M., Landragin, C., Trouchet, A., Bento, S.T., Eisele, A., Foulon, S., Baudre, L., et al. (2021). H3K27me3 is a determinant of chemotolerance in triple-negative breast cancer (*Cancer Biology*).
- Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. *Science* 349, 1483–1489.
- Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer* 12, 323–334.
- Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 37, 471–484.
- Mayran, A., and Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological Chemistry* 293, 13795–13804.
- McCleary, D.F., and Rine, J. (2017). Nutritional Control of Chronological Aging and Heterochromatin in *Saccharomyces cerevisiae*. *Genetics* 205, 1179–1193.
- Miao, K., Lei, J.H., Valecha, M.V., Zhang, A., Xu, J., Wang, L., Lyu, X., Chen, S., Miao, Z., Zhang, X., et al. (2020). NOTCH1 activation compensates BRCA1 deficiency and promotes triple-negative breast cancer formation. *Nat Commun* 11, 3256.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.
- Mittnenzweig, M. (2021). A single-embryo, single-cell time-resolved model for mouse gastrulation. *OPEN ACCESS* 41.
- Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 7, 403–417.
- Montagna, M., Santacatterina, M., Torri, A., Menin, C., Zullato, D., Chieco-Bianchi, L., and D'Andrea, E. (1999). Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. *Oncogene* 18, 4160–4165.
- Morgan, M.A.J., and Shilatifard, A. (2020). Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nature Genetics* 52, 1271–1281.
- Morris, S.A. (2016). Direct lineage reprogramming via pioneer factors; a detour through developmental

gene regulatory networks. *Development* 143, 2696–2705.

Mulligan, A.M., O'Malley, F.P., Ennis, M., Fantus, I.G., and Goodwin, P.J. (2007). Insulin receptor is an independent predictor of a favorable outcome in early stage breast cancer. *Breast Cancer Res Treat* 106, 39–47.

Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* 19, 1218–1227.

Naciri, I., Laisné, M., Ferry, L., Bourmaud, M., Gupta, N., Di Carlo, S., Huna, A., Martin, N., Peduto, L., Bernard, D., et al. (2019). Genetic screens reveal mechanisms for the transcriptional regulation of tissue-specific genes in normal cells and tumors. *Nucleic Acids Research* 47, 3407–3421.

Nandi, A., and Chakrabarti, R. (2020). The many facets of Notch signaling in breast cancer: toward overcoming therapeutic resistance. *Genes Dev.* 34, 1422–1438.

Nguyen, L., W. M. Martens, J., Van Hoeck, A., and Cuppen, E. (2020). Pan-cancer landscape of homologous recombination deficiency. *Nat Commun* 11, 5584.

Nichols, B.A., Oswald, N.W., McMillan, E.A., McGlynn, K., Yan, J., Kim, M.S., Saha, J., Mallipeddi, P.L., LaDuke, S.A., Villalobos, P.A., et al. (2018). *HORMAD1* Is a Negative Prognostic Indicator in Lung Adenocarcinoma and Specifies Resistance to Oxidative and Genotoxic Stress. *Cancer Res.* 78, 6196–6208.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole genome sequences. *Nature* 534, 47–54.

Odunsi, K., Matsuzaki, J., James, S.R., Mhawech-Fauceglia, P., Tsuji, T., Miller, A., Zhang, W., Akers, S.N., Griffiths, E.A., Miliotto, A., et al. (2014). Epigenetic Potentiation of NY-ESO-1 Vaccine Therapy in Human Ovarian Cancer. *Cancer Immunol Res* 2, 37–49.

O'Hagan, H.M., Wang, W., Sen, S., Shields, C.D., Lee, S.S., Zhang, Y.W., Clements, E.G., Cai, Y., Van Neste, L., Easwaran, H., et al. (2011). Oxidative Damage Targets Complexes Containing DNA Methyltransferases, SIRT1 and Polycomb Members to Promoter CpG Islands. *Cancer Cell* 20, 606–619.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* 99, 247–257.

P, D., W, M., B, W., L, M., and B, B. (2014). Oncogenic functions of the cancer-testis antigen SSX on the proliferation, survival, and signaling pathways of cancer cells. *PLoS One* 9, e95136–e95136.

Pal, B., Bouras, T., Shi, W., Vaillant, F., Sheridan, J.M., Fu, N., Breslin, K., Jiang, K., Ritchie, M.E., Young, M., et al. (2013). Global changes in the mammary epigenome are induced by hormonal cues and coordinated by Ezh2. *Cell Rep* 3, 411–426.

Panizza, S., Mendoza, M.A., Berlinger, M., Huang, L., Nicolas, A., Shirahige, K., and Klein, F. (2011). Spo11-Accessory Proteins Link Double-Strand Break Sites to the Chromosome Axis in Early Meiotic Recombination. *Cell* 146, 372–383.

Paredes, J., Correia, A.L., Ribeiro, A.S., Albergaria, A., Milanezi, F., and Schmitt, F.C. (2007). P-cadherin expression in breast cancer: a review. *Breast Cancer Research* 9, 214.

Paret, C., Simon, P., Vormbrock, K., Bender, C., Kölsch, A., Breitkreuz, A., Yildiz, Ö., Omokoko, T., Hubich-Rau, S., Hartmann, C., et al. (2015). CXorf61 is a target for T cell based immunotherapy of triple-negative breast cancer. *Oncotarget* 6, 25356–25367.

Pasini, D., and Di Croce, L. (2016). Emerging roles for Polycomb proteins in cancer. *Current Opinion in Genetics & Development* 36, 50–58.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. *Nature* 556, 463–468.

Pathania, R., Ramachandran, S., Mariappan, G., Thakur, P., Shi, H., Choi, J.-H., Manicassamy, S., Kolhe, R., Prasad, P.D., Sharma, S., et al. (2016). Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. *Cancer Res* 76, 3224–3235.

Pelissier, F.A., Garbe, J.C., Ananthanarayanan, B., Miyano, M., Lin, C., Jokela, T., Kumar, S., Stampfer, M.R., Lorens, J.B., and LaBarge, M.A. (2014). Age-related dysfunction in mechano-transduction impairs differentiation of human mammary epithelial progenitors. *Cell Rep* 7, 1926–1939.

Pellacani, D., Bilenky, M., Kannan, N., Heravi-Moussavi, A., Knapp, D.J.H.F., Gakkhar, S., Moksa, M., Carles, A., Moore, R., Mungall, A.J., et al. (2016). Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell Reports* 17, 2060–2074.

Pellacani, D., Tan, S., Lefort, S., and Eaves, C.J. (2019). Transcriptional regulation of normal human mammary cell heterogeneity and its perturbation in breast cancer. *EMBO J* 38.

Peters, A.H.F.M., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* 107, 323–337.

Petryk, N., Bultmann, S., Bartke, T., and Defossez, P.-A. (2021). Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. *Nucleic Acids Research* 49, 3020–3032.

Pfister, S.X., and Ashworth, A. (2017). Marked for death: targeting epigenetic changes in cancer. *Nature Reviews Drug Discovery* 16, 241–263.

Piccart-Gebhart, M.J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., Smith, I., Gianni, L., Baselga, J., Bell, R., Jackisch, C., et al. (2005). Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *New England Journal of Medicine* 353, 1659–1672.

Poli, V., Fagnocchi, L., Fasciani, A., Cherubini, A., Mazzoleni, S., Ferrillo, S., Miluzio, A., Gaudio, G., Vaira, V., Turdo, A., et al. (2018). MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. *Nat Commun* 9, 1024.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology* 28, 1057–1068.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nature Reviews Molecular Cell Biology* 10, 192–206.

Proia, T.A., Keller, P.J., Gupta, P.B., Klebba, I., Jones, A.D., Sedic, M., Gilmore, H., Tung, N., Naber, S.P., Schnitt, S., et al. (2011). Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* 8, 149–163.

Rajaram, R.D., Buric, D., Caikovski, M., Ayyanan, A., Rougemont, J., Shan, J., Vainio, S.J., Yalcin-Ozuysal, O., and Briskin, C. (2015). Progesterone and Wnt4 control mammary stem cells via myoepithelial crosstalk. *The EMBO Journal* 34, 641–652.

Rajavelu, A., Lungu, C., Emperle, M., Dukatz, M., Bröhm, A., Broche, J., Hanelt, I., Parsa, E., Schiffers, S., Karnik, R., et al. (2018). Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic Acids Research* 46, 9044–9056.

Ramirez, M., Rajaram, S., Steininger, R.J., Osipchuk, D., Roth, M.A., Morinishi, L.S., Evans, L., Ji, W., Hsu, C.-H., Thurley, K., et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun* 7, 10690.

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J., et al. (2008). Transcriptome Analysis of the Normal Human Mammary Cell Commitment and Differentiation Process. *Cell Stem Cell* 3, 109–118.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425–432.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic Reprogramming in Mammalian Development. *Science* 293, 1089–1093.

Renaud, S., Loukinov, D., Alberti, L., Vostrov, A., Kwon, Y.-W., Bosman, F.T., Lobanenkov, V., and Benhattar, J. (2011). BORIS/CTCF-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. *Nucleic Acids Res* 39, 862–873.

- Rhee, I., Jair, K.-W., Yen, R.-W.C., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B., and Schuebel, K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* 404, 1003–1007.
- Rhind, N., and Gilbert, D.M. (2013). DNA Replication Timing. *Cold Spring Harb Perspect Biol* 5, a010132.
- Ribassin-Majed, L., Le-Teuff, G., and Hill, C. (2017). [The frequency of cancer in France: Most recent data and trends]. *Bull Cancer* 104, 20–29.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9–25.
- Rinaldi, V.D., Bolcun-Filas, E., Kogo, H., Kurahashi, H., and Schimenti, J.C. (2017). The DNA Damage Checkpoint Eliminates Mouse Oocytes with Chromosome Synapsis Failure. *Mol. Cell* 67, 1026–1036.e2.
- Robertson, K.D. (2005). DNA methylation and human disease. *Nature Reviews Genetics* 6, 597–610.
- Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. *Nature Reviews Genetics* 1, 11–19.
- Robinson, J.L.L., Holmes, K.A., and Carroll, J.S. (2013). FOXA1 mutations in hormone-dependent cancers. *Front Oncol* 3, 20.
- Rodrigues, M., Manié, É., Popova, T., and Stern, M.H. (2016). BRCAness/défauts de la recombinaison homologue dans les cancers: mécanismes, diagnostic et conséquences thérapeutiques. *Mis e au point* 6.
- Rosario, R., Smith, R.W.P., Adams, I.R., and Anderson, R.A. (2017). RNA immunoprecipitation identifies novel targets of DAZL in human foetal ovary. *Mol. Hum. Reprod.* 23, 177–186.
- Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta* 1839, 627–643.
- Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* 19, 1155–1160.
- Rotili, D., and Mai, A. (2011). Targeting Histone Demethylases. *Genes Cancer* 2, 663–679.
- Rousseaux, S., Bourova-Flin, E., Gao, M., Wang, J., Mi, J.-Q., and Khochbin, S. (2019). Unprogrammed Gene Activation: A Critical Evaluation of Cancer Testis Genes. In *Encyclopedia of Cancer (Third Edition)*, P. Boffetta, and P. Hainaut, eds. (Oxford: Academic Press), pp. 523–530.
- Roussel-Gervais, A., Naciri, I., Kirsh, O., Kasprzyk, L., Velasco, G., Grillo, G., Dubus, P., and Defossez, P.-A. (2017). Loss of the Methyl-CpG-Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes Tumorigenesis. *Cancer Res* 77, 62–73.
- Rudolph, J., and Luger, K. (2020). The secret life of histones. *Science* 369, 33–33.
- Ruff, M., Leyme, A., Cann, F.L., Bonnier, D., Seyec, J.L., Chesnel, F., Fattet, L., Rimokh, R., Baffet, G., and Théret, N. (2015). The Disintegrin and Metalloprotease ADAM12 Is Associated with TGF- β -Induced Epithelial to Mesenchymal Transition. *PLOS ONE* 10, e0139179.
- Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafeina, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 173, 321–337.e10.
- Sanders, M.E., Schuyler, P.A., Dupont, W.D., and Page, D.L. (2005). The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer* 103, 2481–2484.
- Savage, P., Pacis, A., Kuasne, H., Liu, L., Lai, D., Wan, A., Dankner, M., Martinez, C., Muñoz-Ramos, V., Pilon, V., et al. (2020). Chemogenomic profiling of breast cancer patient-derived xenografts reveals targetable vulnerabilities for difficult-to-treat tumors. *Commun Biol* 3.
- Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103, 1412–1417.

- Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326.
- Scully, R., Xie, A., and Nagaraju, G. (2004). Molecular functions of BRCA1 in the DNA damage response. *Cancer Biol Ther* 3, 521–527.
- Semmler, L., Reiter-Brennan, C., and Klein, A. (2019). BRCA1 and Breast Cancer: a Review of the Underlying Mechanisms Resulting in the Tissue-Specific Tumorigenesis in Mutation Carriers. *J Breast Cancer* 22, 1–14.
- Sérandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Métivier, R., et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* 21, 555–565.
- Serizay, J., Dong, Y., Jänes, J., Chesney, M., Cerrato, C., and Ahringer, J. (2020). Tissue-specific profiling reveals distinctive regulatory architectures for ubiquitous, germline and somatic genes (*Genomics*).
- Shahzad, M.M.K., Shin, Y.-H., Matsuo, K., Lu, C., Nishimura, M., Shen, D.-Y., Kang, Y., Hu, W., Mora, E.M., Rodriguez-Aguayo, C., et al. (2013). Biological Significance of HORM-A Domain Containing Protein 1 (HORMAD1) in Epithelial Ovarian Carcinoma. *Cancer Lett* 330, 123–129.
- Shida, A., Futawatari, N., Fukuyama, T., Ichiki, Y., Takahashi, Y., Nishi, Y., Kobayashi, N., Yamazaki, H., and Watanabe, M. (2015). Frequent High Expression of Kita-Kyushu Lung Cancer Antigen-1 (KK-LC-1) in Gastric Cancer. *Anticancer Res.* 35, 3575–3579.
- Shiino, S., Kinoshita, T., Yoshida, M., Jimbo, K., Asaga, S., Takayama, S., and Tsuda, H. (2016). Prognostic Impact of Discordance in Hormone Receptor Status Between Primary and Recurrent Sites in Patients With Recurrent Breast Cancer. *Clin Breast Cancer* 16, e133-140.
- Shin, Y.-H., Choi, Y., Erdin, S.U., Yatsenko, S.A., Kloc, M., Yang, F., Wang, P.J., Meistrich, M.L., and Rajkovic, A. (2010). Hormad1 Mutation Disrupts Synaptonemal Complex Formation, Recombination, and Chromosome Segregation in Mammalian Meiosis. *PLoS Genet* 6.
- Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeligler, S.R., Fried, Y.C., Ainhinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* 513, 115–119.
- Shuvalov, O., Kizenko, A., Petukhov, A., Fedorova, O., Daks, A., Bottrill, A., Snezhkina, A.V., Kudryavtseva, A.V., and Barlev, N. (2020). SEMG1/2 augment energy metabolism of tumor cells. *Cell Death & Disease* 11, 1–14.
- Simpson, A.J.G., Caballero, O.L., Jungbluth, A., Chen, Y.-T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. *Nature Reviews Cancer* 5, 615–625.
- Sinha, A., Agarwal, S., Parashar, D., Verma, A., Saini, S., Jagadish, N., Ansari, A.S., Lohiya, N.K., and Suri, A. (2013). Down regulation of SPAG9 reduces growth and invasive potential of triple-negative breast cancer cells: possible implications in targeted therapy. *Journal of Experimental & Clinical Cancer Research* 32, 69.
- Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274.
- Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8, 272–285.
- Smeenk, G., and van Attikum, H. (2013). The chromatin response to DNA breaks: leaving a mark on genome integrity. *Annu Rev Biochem* 82, 55–80.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics* 14, 204–220.
- Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. *Nature Reviews Molecular Cell Biology* 17, 139–154.
- Sonawane, A.R., Platig, J., Fagny, M., Chen, C.-Y., Paulson, J.N., Lopes-Ramos, C.M., DeMeo, D.L.,

Quackenbush, J., Glass, K., and Kuijjer, M.L. (2017). Understanding Tissue-Specific Gene Regulation. *Cell Rep* 21, 1077–1088.

Soria, G., Polo, S.E., and Almouzni, G. (2012). Prime, repair, restore: the active role of chromatin in the DNA damage response. *Mol Cell* 46, 722–734.

SPF Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Etude à partir des registres des cancers du réseau Francim - Partie 1 : tumeurs solides.

SR, L., IO, E., SJ, S., PH, T., and MJ, van de V. WHO Classification of Tumours of the Breast. Stanzione, M., Baumann, M., Papanikos, F., Dereli, I., Lange, J., Ramlal, A., Tränkner, D., Shibuya, H., de Massy, B., Watanabe, Y., et al. (2016a). Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. *Nat. Cell Biol.* 18, 1208–1220.

Stanzione, M., Baumann, M., Papanikos, F., Dereli, I., Lange, J., Ramlal, A., Tränkner, D., Shibuya, H., de Massy, B., Watanabe, Y., et al. (2016b). Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. *Nat. Cell Biol.* 18, 1208–1220.

Stevanović, S., Pasetto, A., Helman, S.R., Gartner, J.J., Prickett, T.D., Howie, B., Robins, H.S., Robbins, P.F., Klebanoff, C.A., Rosenberg, S.A., et al. (2017). Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* 356, 200–205.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nature Cell Biology* 22, 361–371.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. Stunnenberg, H.G., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., Amit, I., Antonarakis, S.E., Aparicio, S., Arima, T., et al. (2016). The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* 167, 1145–1149.

Su, Z., and Denu, J.M. (2016). Reading the Combinatorial Histone Language. *ACS Chem. Biol.* 11, 564–574.

Tabassum, D.P., and Polyak, K. (2015). Tumorigenesis: it takes a village. *Nature Reviews Cancer* 15, 473–483.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.

Takata, H., Hanafusa, T., Mori, T., Shimura, M., Iida, Y., Ishikawa, K., Yoshikawa, K., Yoshikawa, Y., and Maeshima, K. (2013). Chromatin Compaction Protects Genomic DNA from Radiation Damage. *PLoS One* 8.

Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* 18, 115–126.

Teschendorff, A.E., Miremadi, A., Pinder, S.E., Ellis, I.O., and Caldas, C. (2007). An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biology* 8, R157.

Theunissen, T.W., and Jaenisch, R. (2017). Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development* 144, 4496–4509.

Tlsty, T.D., Romanov, S.R., Kozakiewicz, B.K., Holst, C.R., Haupt, L.M., and Crawford, Y.G. (2001). Loss of chromosomal integrity in human mammary epithelial cells subsequent to escape from senescence. *J Mammary Gland Biol Neoplasia* 6, 235–243.

Tollefsbol, T. (2017). *Handbook of Epigenetics: The New Molecular and Medical Genetics* (Academic Press).

Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin* 65, 87–108.

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ESC differentiation. *Nature* 518, 344–349.

- Uhlén, M., Hallström, B.M., Lindskog, C., Mardinoglu, A., Pontén, F., and Nielsen, J. (2016). Transcriptomics resources of human tissues and organs. *Molecular Systems Biology* 12, 862.
- Vincent-Salomon, A., Benhamo, V., Gravier, E., Rigai, G., Gruel, N., Robin, S., de Rycke, Y., Mariani, O., Pierron, G., Gentien, D., et al. (2013). Genomic Instability: A Stronger Prognostic Marker Than Proliferation for Early Stage Luminal Breast Carcinomas. *PLoS ONE* 8, e76496.
- Visvader, J.E. (2011). Cells of origin in cancer. *Nature* 469, 314–322.
- Visvader, J.E., and Clevers, H. (2016). Tissue-specific designs of stem cell hierarchies. *Nature Cell Biology* 18, 349–355.
- Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* 28, 1143–1158.
- Wang, F., Redding, S., Finkelstein, I.J., Gorman, J., Reichman, D.R., and Greene, E.C. (2013). The promoter search mechanism of *E. coli* RNA polymerase is dominated by three-dimensional diffusion. *Nat Struct Mol Biol* 20, 174–181.
- Wang, H., Xiang, D., Liu, B., He, A., Randle, H.J., Zhang, K.X., Dongre, A., Sachs, N., Clark, A.P., Tao, L., et al. (2019). Inadequate DNA damage repair promotes mammary transdifferentiation leading to BRCA1 breast cancer. *Cell* 178, 135-151.e19.
- Wang, X., Tan, Y., Cao, X., Kim, J.A., Chen, T., Hu, Y., Wexler, M., and Wang, X. (2018). Epigenetic activation of *HORMAD1* in basal-like breast cancer: role in Rucaparib sensitivity. *Oncotarget* 9, 30115–30127.
- Watkins, J., Weekes, D., Shah, V., Gazinska, P., Joshi, S., Sidhu, B., Gillett, C., Pinder, S., Vanoli, F., Jasin, M., et al. (2015). Genomic Complexity Profiling Reveals That *HORMAD1* Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. *Cancer Discov* 5, 488–505.
- Wegiel, B., Bjartell, A., Ekberg, J., Gadaleanu, V., Brunhoff, C., and Persson, J.L. (2005). A role for cyclin A1 in mediating the autocrine expression of vascular endothelial growth factor in prostate cancer. *Oncogene* 24, 6385–6393.
- Wei, X., Chen, F., Xin, K., Wang, Q., Yu, L., Liu, B., and Liu, Q. (2019). Cancer-Testis Antigen Peptide Vaccine for Cancer Immunotherapy: Progress and Prospects. *Transl Oncol* 12, 733–738.
- Wolff, G.L., Kodell, R.L., Moore, S.R., and Cooney, C.A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in *Avy/a* mice. *FASEB J* 12, 949–957.
- Xie, W., Kagiampakis, I., Pan, L., Zhang, Y.W., Murphy, L., Tao, Y., Kong, X., Xia, L., Carvalho, F.L., Sen, S., et al. (2018). DNA methylation patterns separate senescence from transformation potential and indicate cancer risk. *Cancer Cell* 33, 309-321.e5.
- Xu, C.-R., Cole, P.A., Meyers, D.J., Kormish, J., Dent, S., and Zaret, K.S. (2011). Chromatin “Pre-Pattern” and Histone Modifiers in a Fate Choice for Liver and Pancreas. *Science* 332, 963–966.
- Xu, G.L., Bestor, T.H., Bourc’his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402, 187–191.
- Yagi, M., Yamanaka, S., and Yamada, Y. (2017). Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. *Laboratory Investigation* 97, 1133–1141.
- Yamamoto, S., Wu, Z., Russnes, H.G., Takagi, S., Peluffo, G., Vaske, C., Zhao, X., Moen Volla, H.K., Maruyama, R., Ekram, M.B., et al. (2014). *JARID1B* is a luminal lineage-driving oncogene in breast cancer. *Cancer Cell* 25, 762–777.
- Yang, F., Zhou, X., Miao, X., Zhang, T., Hang, X., Tie, R., Liu, N., Tian, F., Wang, F., and Yuan, J. (2014). *MAGEC2*, an epithelial-mesenchymal transition inducer, is associated with breast cancer metastasis. *Breast Cancer Res Treat* 145, 23–32.
- Yao, J., Caballero, O.L., Yung, W.K.A., Weinstein, J.N., Riggins, G.J., Strausberg, R.L., and Zhao, Q. (2014). Tumor subtype-specific cancer-testis antigens as potential biomarkers and immunotherapeutic targets for cancers. *Cancer Immunol Res* 2, 371–379.

- Ye, Q., Kim, D.H., Dereli, I., Rosenberg, S.C., Hagemann, G., Herzog, F., Tóth, A., Cleveland, D.W., and Corbett, K.D. (2017). The AAA+ ATPase TRIP13 remodels HORMA domains through N-terminal engagement and unfolding. *EMBO J.* 36, 2419–2434.
- Yeh, A., Wei, M., Golub, S.B., Yamashiro, D.J., Murty, V.V., and Tycko, B. (2002). Chromosome arm 16q in Wilms tumors: unbalanced chromosomal translocations, loss of heterozygosity, and assessment of the CTCF gene. *Genes Chromosomes Cancer* 35, 156–163.
- Yi, S., Lin, S., Li, Y., Zhao, W., Mills, G.B., and Sahni, N. (2017). Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat Rev Genet* 18, 395–410.
- Yin, L., Duan, J.-J., Bian, X.-W., and Yu, S. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research* 22, 61.
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356.
- Yomtoubian, S., Lee, S.B., Verma, A., Izzo, F., Markowitz, G., Choi, H., Cerchietti, L., Vahdat, L., Brown, K.A., Andreopoulou, E., et al. (2020). Inhibition of EZH2 Catalytic Activity Selectively Targets a Metastatic Subpopulation in Triple-Negative Breast Cancer. *Cell Rep* 30, 755-770.e6.
- Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25, 2227–2241.
- Zaret, K.S., and Mango, S. (2016). Pioneer Transcription Factors, Chromatin Dynamics, and Cell Fate Control. *Curr Opin Genet Dev* 37, 76–81.
- Zaret, K.S., Caravaca, J.M., Tulin, A., and Sekiya, T. (2010). Nuclear Mobility and Mitotic Chromosome Binding Similarities between Pioneer Transcription Factor FoxA and Linker Histone H1. *Cold Spring Harb Symp Quant Biol* 75, 219–226.
- Zhang, M., Tsimelzon, A., Chang, C.-H., Fan, C., Wolff, A., Perou, C.M., Hilsenbeck, S.G., and Rosen, J.M. (2015). Intratumoral heterogeneity in a p53 null mouse model of human breast cancer. *Cancer Discov* 5, 520–533.
- Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics* 12, 7–18.
- Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.
- Zion, E.H., Chandrasekhara, C., and Chen, X. (2020). Asymmetric inheritance of epigenetic states in asymmetrically dividing stem cells. *Current Opinion in Cell Biology* 67, 27–36.
- Zou, M.R., Cao, J., Liu, Z., Huh, S.J., Polyak, K., and Yan, Q. (2014). Histone demethylase jumonji AT-rich interactive domain 1B (JARID1B) controls mammary gland development by regulating key developmental and lineage specification genes. *J Biol Chem* 289, 17620–17633. (2016). *Novel Biomarkers in the Continuum of Breast Cancer* (Cham: Springer International Publishing).