



HAL
open science

Développement de nouveaux tests fonctionnels d'aide à l'interprétation des variants de signification biologique inconnue dans le cadre de prédispositions génétiques au cancer.

Sabine Raad

► To cite this version:

Sabine Raad. Développement de nouveaux tests fonctionnels d'aide à l'interprétation des variants de signification biologique inconnue dans le cadre de prédispositions génétiques au cancer.. Cancer. Normandie Université, 2018. Français. NNT: 2018NORMR079 . tel-04621971

HAL Id: tel-04621971

<https://theses.hal.science/tel-04621971>

Submitted on 24 Jun 2024

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

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



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité Aspects Moléculaires et Cellulaires de la Biologie

Préparée au sein de « l'Université de Rouen Normandie »

Développement de nouveaux tests fonctionnels d'aide à l'interprétation des variants de signification biologique inconnue dans le cadre des prédispositions génétiques au cancer

**Présentée et soutenue par
Sabine RAAD**

Thèse soutenue publiquement le (date de soutenance) devant le jury composé de		
Pr Anne-Paule GIMENEZ-ROQUEPLO	PU-PH / Inserm UMR970 / Hôpital Européen Georges Pompidou, Paris / Université Paris Descartes	Rapporteur
Pr André MEGARBANE	PR / Institut Jérôme Lejeune, Paris / Université Saint Joseph Beyrouth, Liban	Rapporteur
Dr Valérie BONADONA	PH / Unité Clinique d'Oncologie génétique, Centre Léon Bernard, Lyon / Université Claude Bernard Lyon 1	Examineur
Dr Isabelle TOURNIER	MCF / Inserm UMR1245 / Université de Rouen Normandie	Co-encadrante de thèse
Pr Thierry FREBOURG	PU-PH / Inserm UMR1245 / Service de Génétique, CHU de Rouen / Université de Rouen Normandie	Directeur de thèse

Thèse dirigée par le Professeur Thierry FREBOURG Unité Inserm U1245



Cette thèse a été réalisée au sein du laboratoire Inserm UMR1245 dirigé par le Professeur Thierry Frebourg :

Inserm UMR1245 - IRIB
UFR Santé
22 boulevard Gambetta – CS 76183
76183 ROUEN CEDEX 1
FRANCE
Tél : 02 35 14 82 80 - Fax : 02 35 14 82 37
Inserm.u1245@univ-rouen.fr



Financeurs de ces travaux :



La science est assez forte pour envoyer un jour sur la Lune les gens qui sont sur la Terre,
mais comment fera-t-elle *pour ramener sur la Terre les gens qui sont dans la Lune...*

Germaine Beaumont

Remerciements

Je tiens à remercier Madame le Professeur Anne-Paule Gimenez Roqueplo d'avoir accepté d'évaluer mes travaux de thèse malgré son emploi du temps chargé, et Monsieur le Professeur Andre Mégarbané de nous faire l'honneur de juger ce travail et pour l'effort fourni concernant son déplacement depuis le Liban. Merci également à Madame le Docteur Valérie Bonadona d'avoir accepté d'examiner mes travaux de thèse.

J'aimerais aussi remercier les membres de mon comité de suivi individuel de thèse : Dr. Béatrice Parfait, Dr. Sylvie Mazoyer, Dr. Myriam Bernaudin et Dr Inès Boukhalfa qui m'ont fait l'honneur de participer à mes deux comités de suivi de thèse. Vos conseils ont été plus que bénéfiques.

Je tiens à remercier très sincèrement le Professeur Thierry Frebourg de m'avoir accueillie au sein de son laboratoire de recherche depuis mon Master. Merci d'avoir dirigé mes travaux de thèse et de m'avoir fait confiance tout au long de ces années, merci pour votre investissement et vos nombreux conseils que vous m'avez prodigués dans la réalisation de ce projet et surtout ces derniers mois, merci de penser à notre avenir professionnel et d'être présent pour nous rassurer, merci pour votre investissement personnel votre support à mon égard et votre gentillesse, merci également de nous faire partager votre enthousiasme pour la génétique humaine, vos connaissances et votre passion m'ont toujours émerveillées et ont été une vraie stimulation.

Isabelle, ce petit paragraphe ne peut pas résumer tous les mots et les remerciements que je souhaite t'adresser. Depuis mon premier jour de stage il y a 5 ans, t'as cru en moi et grâce à la chance que tu m'aies offerte, aujourd'hui je suis là. Sans toi rien n'aurait été possible et je n'oublierai pas ton grand investissement et effort pour que je puisse réaliser ma thèse au sein de ton équipe. Merci pour tes conseils, pour toutes les heures de discussions scientifiques que tu m'apportes chaque jour, merci d'avoir pris le temps pour m'expliquer tout, de me former aux manip, de me faire découvrir le monde du NGS et pour le temps que t'as passé pour corriger le manuscrit malgré ton emploi de temps surchargé. J'ai eu la chance d'être ta première thésarde (fille) pour profiter de tes connaissances énormes, ta rigueur scientifique, tes nombreuses idées et ta passion pour la recherche. Merci de m'avoir épaulée tout au long de ces années et même pendant tes congés maternités où tu as toujours été disponible. Merci de m'avoir encouragée et supportée pour que je puisse construire ma carrière professionnelle, d'avoir accepté mes souhaits surtout concernant l'enseignement et mes futurs projets. Chef, tu m'as comblée par ta gentillesse, ta bonne humeur au quotidien, tes qualités humaines qui m'ont fait grandir et ton soutien inépuisable. Ta présence et ton amour m'ont constitué une motivation et une force pour réaliser ces travaux et surmonter l'absence de mes proches. Merci d'avoir toujours répondu à mes questions et de me divertir quand j'en avais besoin, pour ton écoute et tes conseils tant d'un point de vue professionnel que personnel ! Merci de m'avoir fait rencontrer ta belle famille, et merci pour les belles sorties qui m'ont permis de découvrir la Normandie. Isabelle, tu es une encadrante formidable avec qui j'ai adoré travailler. Je me considère chanceuse d'avoir fait ta connaissance. Je regrette que l'histoire s'arrête là et pour sûr tu vas trop me manquer. J'espère que nos routes scientifiques et personnelles se recroiseront souvent... Et maintenant que tu m'as permis d'en arriver là c'est le moment de t'emmener explorer le Liban.

Gaëlle, merci pour tes précieux conseils et ton aide tout au long de ma thèse. Ton investissement a grandement contribué à l'avancement de mes travaux. Jean-Michel, merci de m'avoir fait découvrir le monde des levures et de m'avoir consacré du temps avant ton départ. Travailler avec vous fut un vrai plaisir.

Nathalie, merci pour ta gentillesse, ta bonne humeur au quotidien, ta disponibilité, d'être là pour nous au jour le jour et d'être toujours prête à nous rendre service, merci d'illuminer nos journées et surtout nous écouter et nous encourager.

Cher Pierre, aussi je n'ai pas beaucoup de mots pour te dire merci et je ne saurai pas t'exprimer toute ma reconnaissance. Je voudrais te remercier du fond du cœur pour avoir été comme mon grand frère au laboratoire et en dehors pendant toutes ces années. Ta rencontre était exceptionnelle tant sur le plan scientifique que sur le plan humain. Tu m'as transmis toutes tes connaissances, ton soutien et ta disponibilité de tous les instants. Merci pour ta bonne humeur, ton support pendant mes moments difficiles, mes moments de maladies, et même dans mes déménagements. J'emporte avec moi le souvenir de tous ces moments partagés. Grâce à tes théories (de pratique) tout semble beaucoup plus facile. Merci pour les fous rires, ta magie, ta gentillesse et pour toutes ces petites choses qui font que les journées sont plus agréables. Tu es une personne formidable chef et je t'aime

beaucoup. Merci également à toi Inès, pour ton engagement dans mon CSI, pour tes conseils et pour ta grande gentillesse, merci à tes parents pour votre précieux soutien.

Edvwige, je te remercie du fond du cœur pour ta disponibilité, ton aide sur ce projet et de m'avoir initiée aux manips. C'est tellement agréable de travailler avec toi ! merci de me faire confiance et de m'avoir léguée tes souris pendant ton absence, merci pour ta belle compagnie, ta sympathie, ton encouragement pendant les congrès et ton amitié. Je te souhaite tout le meilleur pour toi et pour ta famille avec ton petit blond que j'ai hâte de voir. J'espère que nous garderons contact très longtemps.

Alex, mon ami, on est déjà là ! Nous avons commencé cette aventure ensemble depuis 5 ans et nous avons passé de très bons moments. Durant ces années tu as toujours été sympa et gentil, tu m'as formé à la cytométrie et tu m'as trop aidé pour la culture cellulaire et t'étais toujours là pour moi malgré tes journées chargées. Grâce à toi j'aime plus que jamais le foot (mais je ne soutiendrai jamais le PSG désolée), j'aime Julien Doré et oui je comprends mieux l'immunologie. Je suis vraiment heureuse de croiser ton chemin, tu vas me manquer énormément mais je sais qu'on restera en contact malgré les distances, et comme t'as dit on se verra au Liban en 2020 pour fêter (en espérant que ton pronostic soit une réalité) et peut être tu seras le coach de l'équipe Libanaise ! Enfin, je te souhaite une belle réussite parisienne et plein de bonnes choses que tu mérites tonton Alex et que la force soit avec toi le 14.

Sacré Marinette, merci pour ta folie partagée, ta joie de vivre et pour cette vie que tu as apportée dans le bureau. Tu m'auras appris plein de truc sur les chevaux et grâce à toi je connais beaucoup de chansons françaises et des blagues rigolotes. Je n'ai qu'un seul souhait : Te voir en robe et j'espère que le jour J se rapproche :P Merci pour ton soutien humain intense pendant toute cette période, pour être toujours présente, toujours à l'écoute et pour les beaux souvenirs.

Raphael, un merci particulier pour ta patience sur l'analyse bioinformatique du RNA-Seq que grâce à tes schémas je comprends beaucoup mieux. Je te préviens que je ne vais pas croire ta blague d'alignement !! Sophie, un grand merci pour ta grande gentillesse au quotidien et ton efficacité. Steve, merci pour ton aide, ta sympathie et pour la bonne ambiance que t'as créée ces derniers mois.

Un merci sincère et chaleureux pour toi Céline, c'était un plaisir de travailler avec toi et merci pour ton aide sur le plan personnel.

Marion, merci pour ta contribution à ce travail, merci pour ta gentillesse et ta patience, c'était sympa de travailler avec toi.

Je souhaite également bonne chance aux petits nouveaux du laboratoire : Laetitia, merci pour ta belle compagnie durant les congrès, tes belles vidéos et ta gentillesse, Camille bonne continuité et bon courage.

Une pensée particulière à ceux que je ne vois plus mais qui ont passé leur thèse dans le laboratoire Estelle, Kilan, et Hafid.

Un grand merci à toute l'équipe de l'unité Inserm U1245 pour leur soutien tout au long de ce parcours et pour leur aide précieuse : Xana, Pascaline, Gaël, Magalie, Anne, Chirine, Jean-Baptiste, Françoise, Myriam, Aurélie, Hélène, Ludivine, Olivier, Sébastien, Laëtitia, Anne-Claire, Stéphane, Camille, Rose-Monde, Thomas, Kevin, François, Maud, Pauline et Maelle. Merci pour votre aide dans les manips et nos discussions et merci d'avoir toujours répondu à mes questions. Merci à toute l'équipe du 2^{ème} étage : Stéphanie, Patricia et également Nathalie, Virginie, Sandrine, Stéphanie, Jackie, Emilie, Séverine, Gwen, merci à vous pour votre gentillesse et votre aide dans les manips.

Gaëtan, merci pour ton aide et ton investissement dans toutes les manips de ce projet.

Ammoura, ces 5 années sont gravées à jamais dans mon cœur et c'est avec beaucoup de nostalgie que je repense à tous les bons moments passés. Ta présence m'a beaucoup réconfortée dans les moments difficiles. Un grand merci pour ton soutien, ta sympathie, ta disponibilité et ta présence tout au long de ces années. C'était un plaisir

immense que j'avais chaque matin à te retrouver. En commençant une thèse, je ne me doutais pas que je repartirais de ce laboratoire avec un véritable ami, plus précieux que des résultats scientifiques.

Roky, merci pour toutes ces heures passées au téléphone, merci pour tes conseils, ta tendresse, ta présence durant les moments durs pour me soulager et me donner confiance, pour tous les moments que nous avons partagés durant ces 9 années des plus difficiles aux meilleurs, en espérant pouvoir en partager encore tant d'autres. La distance, les années, et les changements de cap ne nous ont pas séparés, et j'en suis heureuse. T'avoir à mes côtés est un cadeau si exceptionnel, que j'ai besoin de toute ma vie pour te garder. Je t'aime énormément et je te souhaite bonne chance pour ta thèse.

Wafaa, merci pour tous ces bons moments de détente nécessaires pour recharger les batteries, merci d'avoir été là ces 5 années, et d'être mon soutien dans les moments difficiles et pour tout ce que nous avons partagé. Tous ces moments sont devenus inoubliables grâce à toi. Je t'aime beaucoup.

Dudi, Maya et Zaza merci pour votre amitié au-delà des milliers de kms. Merci de rentrer du Liban pour partager avec moi ce jour tant attendu. Je suis fière de vous avoir à mes côtés pour ce jour particulier et je vous aime énormément. Tuta, merci pour ta présence et ton amitié love uu ...

Fedo merci d'avoir partagé les moments de galères et les moments de joie et d'être à mes côtés quand j'ai besoin de toi. Love u fidz.

Rimo merci pour ta présence depuis le premier jour, t'étais comme ma maman sur Rouen et j'ai hâte de fêter ton mariage soon. Je t'aime beaucoup. Zweze un grand merci pour toi aussi et tes blagues qui m'ont amélioré le moral. Bonne chance pour votre nouvelle vie.

Un grand merci à Najo, Naila, Youssef, Mahmoud, Léa, Grace, Kawthar, Ismo, Imane, Ali, Hussein, Batoul et Jamila. Bonne chance pour la suite. Merci à Razo, Rimo, Saluha, Batta, Fifi, Roro et Layla.

A ma cousine Rania, son mari Yannick et leurs enfants, merci pour votre présence, je vous aime beaucoup.

A mes sœurs Nermine et Diala, merci de vous avoir à mes côtés, dans ma tête et dans mon cœur, merci de votre soutien au quotidien, pour avoir supporté ma mauvaise humeur dans les jours difficiles, votre amour sans condition m'a permis de me conduire jusqu'ici.

A mon papa, pour ton amour et tes encouragements, qui m'ont permis de ne jamais abandonner et de réaliser mon rêve. Pendant tout ce temps, tu étais là ! Et aujourd'hui, je ne peux imaginer que j'en serais arrivé là sans toi. Mama, merci de m'avoir soutenue dans mes choix, merci de ta présence dans ma vie et ton amour éternel. Cette thèse est un peu la vôtre.

Parfois notre lumière s'éteint, puis elle est rallumée par un autre être humain. Chacun de nous doit de sincères remerciements à ceux qui ont ravivé leur flamme.

Albert Schweitzer

RESUME

L'identification des mutations constitutionnelles à l'origine d'une prédisposition génétique au cancer est essentielle à la prise en charge médicale des patients et de leurs familles. Depuis l'implémentation des technologies de séquençage à haut-débit dans les laboratoires diagnostiques, le principal défi n'est plus la détection des variations génétiques mais leur interprétation et leur classification. La question de l'interprétation de la variation est particulièrement cruciale lorsqu'elle conditionne la stratégie thérapeutique. Ainsi, il est essentiel de disposer de tests simples adaptables en routine diagnostique pour faciliter l'interprétation des variations génétiques. Dans ce contexte, nous avons utilisé un test fonctionnel développé par notre équipe pour classer des variations dans le gène TP53 à l'origine du syndrome de Li-Fraumeni et pour appréhender la corrélation génotype - phénotype chez les patients LFS. Dans un deuxième temps, nous avons évalué la pertinence d'une approche multi-omique (RNA-Seq et métabolomique) pour discriminer les cellules sauvages des cellules avec mutation hétérozygote du gène TP53 ou des gènes BRCA impliqués dans la prédisposition génétique aux cancers du sein et de l'ovaire. Sur la base des données de transcriptome, un modèle mathématique a été développé pour détecter les variants correspondant à des mutations délétères. Nous avons ensuite sélectionné les biomarqueurs les plus discriminants pour les intégrer dans un test fonctionnel de RT-MLPA dédié à la voie p53. Nous avons enfin adapté cet essai pour qu'il soit réalisable sur une simple prise de sang, sans immortalisation des lymphocytes du patient.

Mots clés : Prédispositions génétiques au cancer, Approche multi-omique, Métabolomique, RNA-seq, Tests fonctionnels, TP53, BRCA

SUMMARY

The identification of the constitutional mutation responsible for a genetic predisposition to cancer is essential to the clinical management of the patient and its relatives. With the implementation of high-throughput sequencing to the diagnostic routine of these pathologies, the challenge no longer lies within the detection of alterations but in their biological and clinical interpretation. While specific treatments are emerging, simple functional assays to help with the interpretation of the detected variants are needed. In this context, we used a functional test developed by our team to classify variations in the TP53 gene responsible for Li-Fraumeni syndrome and to understand the genotype-phenotype correlation in LFS patients. On the other hand, we assessed the relevance of a multi-omic approach (RNA-Seq and metabolomics) to discriminate wild-type cells from cells with a deleterious heterozygous mutation in TP53 or in the BRCA genes implicated in genetic predisposition to breast and ovarian cancers. Based on the transcriptomic data, a mathematical model has been developed to detect variants corresponding to deleterious mutations. Then we selected the most discriminating biomarkers and integrated them into a RT-MLPA functional assay dedicated to the p53 pathway. We finally adapted this test to be feasible on a simple blood test, without immortalization of the patient's lymphocytes.

Keywords: Genetic cancer predispositions, Multi-omic approach, RNA-Seq, Metabolomics, Functional assays, TP53, BRCA

TABLE DES MATIERES

TABLE DES MATIERES	14
LISTE DES FIGURES	16
LISTE DES TABLEAUX	18
LISTE DES FIGURES ANNEXES	18
LISTE DES TABLEAUX ANNEXES	18
PUBLICATION ANNEXE	18
LISTE DES ABBREVIATIONS	19
INTRODUCTION	22
CHAPITRE I	23
LES PREDISPOSITIONS HEREDITAIRES AU CANCER	23
I - Le cancer, une maladie génétique.....	24
II - Formes mendéliennes de cancer	30
III- Diagnostic des formes héréditaires de cancer	33
CHAPITRE II	38
PLACE DU SEQUENÇAGE HAUT-DEBIT DANS LE DIAGNOSTIC DES PREDISPOSITIONS HEREDITAIRES AU CANCER	38
I- <i>De la découverte de la structure de l'ADN au séquençage du génome humain</i>	39
II- La révolution du séquençage à haut-débit	41
III- Principe du séquençage NGS.....	44
IV- Diagnostic des formes héréditaires de cancer et NGS.....	56
CHAPITRE III	62
INTERPRETATION DES VARIANTES DANS LE CADRE DES FORMES MENDELIENNES DE CANCER	62
I- Classification des variants	63
II- <i>Approches d'interprétation des variants</i>	65
III- <i>Le challenge d'interprétation des variants à l'ère du séquençage tout génome</i>	76
CHAPITRE IV	78
TESTS FONCTIONNELS DANS LE CADRE DES PREDISPOSITIONS HEREDITAIRES	78
I- <i>Cancer héréditaire du sein et de l'ovaire</i>	79
II- Le syndrome de Li-Fraumeni (LFS)	102
OBJECTIF 1	129
ÉTUDE DE L'IMPACT DES MUTATIONS CONSTITUTIONNELLES DE TP53 SUR LA LIAISON A L'ADN DE P53 ET SUR LA REPONSE TRANSCRIPTIONNELLE AUX LESIONS DE L'ADN	129
I- CONTEXTE SCIENTIFIQUE.....	130
II- RESULTATS	132
III- DISCUSSION	151
OBJECTIF 2	153
RECHERCHE DE NOUVEAUX BIOMARQUEURS INDICATEURS DE LA PRESENCE	

<i>D'UNE ALTERATION CONSTITUTIONNELLE DES GENES SUPPRESSEURS DE TUMEURS TP53 ET BRCA PAR APPROCHE MULTI-OMIQUE.....</i>	153
I- CONTEXTE SCIENTIFIQUE.....	154
II- MATERIELS ET METHODES	158
III- RESULTATS	174
IV- DISCUSSION	211
CONCLUSION.....	215
REFERENCES BIBLIOGRAPHIQUES.....	221
ANNEXE	247

LISTE DES FIGURES

FIGURE 1 : LE MOMENT D'INTERVENTION DES DIFFERENTS SYSTEMES DE REPARATION DE L'ADN AU COURS DU CYCLE CELLULAIRE (D'APRES GEE ET AL., 2018).....	27
FIGURE 2 : CAPACITES ACQUISES PAR LES CELLULES CANCEREUSES.....	28
FIGURE 3 : PHYLOGENIE GENETIQUE DES TUMEURS (ADAPTEE DE BRAEKELEER ET AL., 2015).....	29
FIGURE 4 : SCHEMA PRESENTANT LE PRINCIPE DE FONCTIONNEMENT DES INHIBITEURS DE PARP (FIGURE ADAPTEE DE TURK & WISINSKI, 2018).....	37
FIGURE 5 : COMPARAISON ENTRE LE SEQUENÇAGE TRADITIONNEL DE SANGER ET LA TECHNOLOGIE DE SEQUENÇAGE DE NOUVELLE GENERATION (D'APRES PRICE ET AL., 2018).....	42
FIGURE 6 : GRAPHIQUE ILLUSTRANT LA CAPACITE DE SEQUENÇAGE MONDIALE EN TERABASES EN FONCTION DU TEMPS (DONNEES ISSUES DU SRA (<i>SEQUENCE READ ARCHIVE</i>) A LA DATE DU 2 AOUT 2018, HTTPS://TRACE.NCBI.NLM.NIH.GOV/TRACES/SRA/SRA.CGI).....	43
FIGURE 7 : GRAPHE ILLUSTRANT L'EVOLUTION DU COUT DE SEQUENÇAGE D'UN GENOME COMPLET EN FONCTION DU TEMPS (HTTPS://WWW.GENOME.GOV/SEQUENCINGCOSTSDATA/ , A LA DATE DU 25 AVRIL 2018) (DONNEES ISSUES DU <i>NATIONAL HUMAN GENOME RESEARCH, INSTITUTE GENOME SEQUENCING PROGRAM GSP</i>).....	43
FIGURE 8 : ENRICHISSEMENT PAR CAPTURE EN PHASE LIQUIDE DES EN REGIONS D'INTERET – TECHNOLOGIE SURESELECT (AGILENT).....	46
FIGURE 9 : REPRESENTATION SCHEMATIQUE DU PRINCIPE D'ENRICHISSEMENT PAR PCR MULTIPLEXE (SOURCE : HTTPS://WWW.AGILENT.COM).....	47
FIGURE 10 : SCHEMA MONTRANT LES DEUX METHODES D'AMPLIFICATION CLONALE DES LIBRAIRES (D'APRES BAHASSI & STAMBROOK, 2014).....	49
FIGURE 11 : APERÇU DES TECHNOLOGIES DE SEQUENÇAGE DE DEUXIEME GENERATION (D'APRES WONG, 2013)....	52
FIGURE 12 : CLASSEMENT DES VARIANTS DETECTES SELON LES RECOMMANDATIONS DE L'ACMG.....	64
FIGURE 13 : TAILLE DES BASES DE DONNEES DE POPULATION GENERALE EN NOMBRE D'EXOMES ANALYSES ET REPARTITION DES INDIVIDUS SELON LEUR ORIGINE ETHNIQUE.....	67
FIGURE 14 : LES DIFFERENTES ETAPES POUR LA REALISATION DU TEST MINIGENE ANALYSANT L'EFFET DES VARIANTES SUR L'EPISSAGE.....	74
FIGURE 15 : DOUBLE LETALITE SYNTHETIQUE DANS LES CELLULES <i>BRCA</i> DEFICIENTES (D'APRES SULLIVAN-REED ET AL., 2018).....	82
FIGURE 16 : STRUCTURE DU GENE <i>BRC1</i> ET DOMAINES PROTEIQUES IDENTIFIES DE LA PROTEINE <i>BRC1</i> (D'APRES HOLLIS ET AL., 2017).....	83
FIGURE 17 : STRUCTURE DU GENE <i>BRC2</i> ET DOMAINES PROTEIQUES IDENTIFIES DE LA PROTEINE <i>BRC2</i> (D'APRES HOLLIS ET AL., 2017).....	85
FIGURE 18 : LA REPARATION DES CASSURES DOUBLE BRIN DE L'ADN PAR RECOMBINAISON HOMOLOGUE (HR) (D'APRES CHEN ET AL., 2018).....	87
FIGURE 19 : PRINCIPE DE L'ESSAI HDR.....	92
FIGURE 20 : PRINCIPE DU TEST MULTIPLEXE DE HDR (D'APRES STARITA ET AL., 2018).....	93
FIGURE 21: FORMATION DE <i>FOC1</i> -H2AX EN REPONSE AUX CASSURES DOUBLE BRIN DE L'ADN.....	95
FIGURE 22: FORMATION DES <i>FOC1</i> RAD51 EN REPONSE AUX CASSURES DOUBLE BRIN DE L'ADN.....	96
FIGURE 23 : PRINCIPE DE L'ESSAI FONCTIONNEL BASE SUR L'UTILISATION DES CELLULES SOUCHES EMBRYONNAIRES MODIFIEES DE SOURIS.....	99
FIGURE 24 : SCHEMA ILLUSTRANT LES ETAPES DU TEST A HAUT-DEBIT (D'APRES FINDLAY ET AL., 2018).....	100
FIGURE 25 : DOMAINES DE LA PROTEINE P53 HUMAINE (D'APRES TANAKA ET AL., 2018).....	108
FIGURE 26 : REPRESENTATION DE LA STRUCTURE CRISTALLOGRAPHIQUE DU DOMAINE DE TETRAMERISATION DE P53 DANS 3 ORIENTATIONS DIFFERENTES (D'APRES TANAKA ET AL., 2018).....	108
FIGURE 27 : REGULATION DU NIVEAU ET DE L'ACTIVITE DE LA PROTEINE P53 DANS LES CELLULES NON STRESSEES PAR RAPPORT AUX CELLULES STRESSEES (D'APRES AUBREY ET AL., 2018).....	110
FIGURE 28 : SEQUENCE CONSENSUS D'ADN DES ELEMENTS DE REPONSE A P53 (D'APRES TANAKA ET AL., 2018)..	111
FIGURE 29: ROLES DE LA PROTEINE P53 (D'APRES MELLO & ATTARDI, 2018).....	117
FIGURE 30 : PROPORTION DES DIFFERENTS TYPES DE MUTATIONS CONSTITUTIONNELLES DU GENE <i>TP53</i> (D'APRES <i>IARC TP53 DATABASE, R19, AOUT 2018</i>).....	118
FIGURE 31: DISTRIBUTION DES MUTATIONS CONSTITUTIONNELLES DU GENE <i>TP53</i> (D'APRES <i>IARC TP53 DATABASE, R19, AOUT 2018</i>).....	119
FIGURE 32: IMPACT FONCTIONNEL DES MUTATIONS DU GENE <i>TP53</i> (D'APRES AUBREY ET AL., 2018).....	122
FIGURE 33 : PRINCIPE DE LA TECHNIQUE FASAY (D'APRES FLAMAN ET AL., 1995).....	125

FIGURE 34: DEPLETION DES ARN RIBOSOMIQUES	164
FIGURE 35: PREPARATION DES LIBRAIRIES EN VUE D'UNE ANALYSE RNA-SEQ	165
FIGURE 36: PIPELINE BIOINFORMATIQUE AURIGA DEDIE A L'ANALYSE D'EXPRESSION DIFFERENTIELLE	167
FIGURE 37: ANALYSE OPLS-DA DES DONNEES DE METABOLOMIQUE OBTENUES POUR LES LIGNEES MUTEES POUR <i>TP53</i> ET LES LIGNEES TEMOINS.....	177
FIGURE 38 : ANALYSE OPLS-DA DES DONNEES DE METABOLOMIQUE OBTENUES POUR LES LIGNEES MUTEES POUR L'UN DES GENES <i>BRC1</i> ET LES LIGNEES TEMOINS.....	178
FIGURE 39 : RESULTATS DE L'ALIGNEMENT SUR LE GENOME DE REFERENCE HUMAIN HG19	182
FIGURE 40 : CLUSTERING HIERARCHIQUE (NON SUPERVISE) DES DONNEES DE RNA-SEQ OBTENUES POUR LES LIGNEES TEMOINS ET LES LIGNEES MUTEES <i>TP53</i> AVEC ET SANS TRAITEMENT A LA DOXORUBICINE	184
FIGURE 41 : ANALYSE D'ONTOLOGIE REALISEE SUR LES GENES DIFFERENTIELLEMENT EXPRIMES EN CONDITIONS BASALES ENTRE LES LIGNEES MUTEES POUR <i>TP53</i> ET LES LIGNEES TEMOINS	186
FIGURE 42: ANALYSE OPLS-DA DES DONNEES DE RNA-SEQ OBTENUES POUR LES LIGNEES MUTEES POUR LE GENE <i>TP53</i> ET LES LIGNEES TEMOINS A L'ETAT BASAL	187
FIGURE 43: ANALYSE D'ONTOLOGIE REALISEE SUR LES GENES DIFFERENTIELLEMENT EXPRIMES EN PRESENCE DE LA DOXORUBICINE ENTRE LES LIGNEES MUTEES POUR <i>TP53</i> ET LES LIGNEES TEMOINS.....	188
FIGURE 44 : ANALYSE OPLS-DA DES DONNEES DE RNA-SEQ OBTENUES POUR LES LIGNEES MUTEES POUR LE GENE <i>TP53</i> ET LES LIGNEES TEMOINS EN PRESENCE DE LA DOXORUBICINE	189
FIGURE 45 : ANALYSE DE CLUSTERING HIERARCHIQUE (NON SUPERVISEE) DES DONNEES DE RNA-SEQ OBTENUES POUR TOUTES LES LIGNEES TEMOINS ET LES LIGNEES MUTEES <i>BRC1</i> APRES TRAITEMENT A LA BLEOMYCINE	190
FIGURE 46: ANALYSE D'ONTOLOGIE REALISEE SUR LES GENES DIFFERENTIELLEMENT EXPRIMES EN PRESENCE DE LA BLEOMYCINE ENTRE LES LIGNEES MUTEES POUR L'UN DES GENES <i>BRC1</i> ET LES LIGNEES TEMOINS	191
FIGURE 47 : ANALYSE OPLS-DA DES DONNEES DE RNA-SEQ OBTENUES EN PRESENCE DE LA BLEOMYCINE POUR LES LIGNEES MUTEES DANS L'UN DES GENES <i>BRC1</i> ET LES LIGNEES TEMOINS	192
FIGURE 48 : ANALYSE PAR RT-QMSPF DES NIVEAUX RELATIFS D'EXPRESSION A L'ETAT BASAL DES GENES SELECTIONNES PAR RNA-SEQ DANS LA LIGNEE IMMORTALISEE R248W DU GENE <i>TP53</i>	196
FIGURE 49 : ANALYSE PAR RT-QMSPF DES NIVEAUX RELATIFS D'EXPRESSION A L'ETAT BASAL DES GENES SELECTIONNES PAR RNA-SEQ DANS LA LIGNEE IMMORTALISEE G302R DU GENE <i>TP53</i>	197
FIGURE 50 : PRINCIPE DE LA TECHNIQUE DE RT-MLPA (D'APRES HTTPS://WWW.MLPA.COM).....	200
FIGURE 51 : ÉVALUATION DU SCORE DE FONCTIONNALITE DE P53 GRACE A L'ESSAI FONCTIONNEL DE RT-MLPA	202
FIGURE 52 : ÉVALUATION DU SCORE DE FONCTIONNALITE DE P53 GRACE A L'ESSAI FONCTIONNEL DE RT-MLPA	203
FIGURE 53 : ÉVALUATION DU SCORE DE FONCTIONNALITE DE P53 GRACE A L'ESSAI FONCTIONNEL DE RT-QMPSF	204
FIGURE 54 : ÉVALUATION DU SCORE DE FONCTIONNALITE DE P53 GRACE A L'ESSAI FONCTIONNEL DE RT-MLPA DES PBMC DES DEUX TEMOINS POUR UN PRELEVEMENT A J+2	208
FIGURE 55: ÉVALUATION DU SCORE DE FONCTIONNALITE DE P53 GRACE A L'ESSAI FONCTIONNEL DE RT-QMPSF DES PBMC DES DEUX TEMOINS POUR UN PRELEVEMENT A J+2	209
FIGURE 56 : SUIVI DES ETAPES DU NOUVEAU TEST FONCTIONNEL <i>TP53</i> REALISE DIRECTEMENT SUR UN PRELEVEMENT SANGUIN	210
FIGURE 57 : SUIVI DES ETAPES DU NOUVEAU TEST FONCTIONNEL <i>TP53</i> REALISE DIRECTEMENT SUR UN PRELEVEMENT SANGUIN	210

LISTE DES TABLEAUX

TABLEAU 1: DESCRIPTION DES CRITERES D'AMSTERDAM I ET II ET DES REVISED BETHESDA GUIDELINES (D'APRES PRADERE ET AL., 2017)	31
TABLEAU 2 : LES PRINCIPALES PREDISPOSITIONS GENETIQUES AU CANCER ET LES PRINCIPAUX GENES ASSOCIES A CES PREDISPOSITIONS (DONNEES ISSUES DE L'INCA, INSTITUT NATIONAL DU CANCER, 2018, HTTPS://WWW.E-CANCER.FR/)	32
TABLEAU 3: PRESENTATION DES QUATRE PORTES D'ENTREE CLINIQUES DU SYNDROME DE LI-FRAUMENI D'APRES LES CRITERES DE CHOMPRET ACTUALISES EN 2015 (BOUGEARD ET AL., 2015)	104
TABLEAU 4 : LIGNEES DE LYMPHOCYTES IMMORTALISEES DES PATIENTS UTILISEES POUR LES DIFFERENTES APPROCHES EXPERIMENTALES	159
TABLEAU 5: LES SCORES D'INDUCTION DE P53 OBTENUS EN TEST FONCTIONNEL DE RT-QMPSF APRES TRAITEMENT AVEC LA DOXORUBICINE OU LA BLEOMYCINE POUR TOUTES LES LIGNEES ANALYSEES EN RNA-SEQ	175
TABLEAU 6: CONTROLE QUALITE DES DONNEES DE RNA-SEQ	180
TABLEAU 7 : FONCTIONS DES 6 GENES SELECTIONNES SUR LA BASE DE L'ANALYSE D'EXPRESSION DIFFERENTIELLE DESEQ2 DES DONNEES DE RNA-SEQ ET ANALYSES EN RT-QMPSF	194
TABLEAU 8 : FONCTIONS DES 10 GENES SELECTIONNES SUR LA BASE DE L'ANALYSE D'EXPRESSION DIFFERENTIELLE DESEQ2 DES DONNEES DE RNA-SEQ ET ANALYSES EN RT-MLPA	198
TABLEAU 9: LES SCORES OBTENUS PAR LES TECHNIQUES DE RT-MLPA ET RT-QMPSF POUR LES 4 LIGNEES TEMOINS ET POUR LES 35 LIGNEES IMMORTALISEES MUTEES POUR LE GENE TP53	205

LISTE DES FIGURES ANNEXES

FIGURE ANNEXE 1 : NOMBRE DE READS OBTENUS POUR LES 6 GENES EN RT-QMPSF A L'ISSU DE L'ANALYSE DE RNA-SEQ CHEZ LES TEMOINS ET CHEZ LES MUTANTS TP53 SANS TRAITEMENT (NT) OU AVEC TRAITEMENT A LA DOXORUBICINE (DOXO)	260
FIGURE ANNEXE 2 : NOMBRE DE READS OBTENUS POUR LES 10 GENES EN RT-MLPA A L'ISSU DE L'ANALYSE DE RNA-SEQ CHEZ LES TEMOINS ET CHEZ LES MUTANTS TP53 SANS TRAITEMENT (NT) OU AVEC TRAITEMENT A LA DOXORUBICINE (DOXO)	262
FIGURE ANNEXE 3 : NOMBRE DE READS OBTENUS POUR LES 3 GENES CONTROLES EN RT-MLPA A L'ISSU DE L'ANALYSE DE RNA-SEQ CHEZ LES TEMOINS ET CHEZ LES MUTANTS TP53 SANS TRAITEMENT (NT) OU AVEC TRAITEMENT A LA DOXORUBICINE (DOXO)	262

LISTE DES TABLEAUX ANNEXES

TABLEAU ANNEXE 1 : SEQUENCES DES SONDAS UTILISEES EN RT-MLPA	257
---	-----

PUBLICATION ANNEXE

ARTICLE ANNEXE 1 : CONTRIBUTION OF GENOTOXIC ANTICANCER TREATMENTS TO THE DEVELOPMENT OF MULTIPLE PRIMARY TUMOURS IN THE CONTEXT OF GERMLINE TP53 MUTATIONS	247
--	-----

LISTE DES ABBREVIATIONS

ACP	<i>Principal Component Analysis</i>
ADN	<i>Acide DésoxyriboNucléique</i>
AFP	<i>Alpha Fetoprotein</i>
AIR	<i>Ribosyl-Amino-Imidazole</i>
AMM	<i>Autorisation de Mise sur le Marché</i>
AR	<i>Androgen Receptors</i>
ARNg	<i>ARN guide</i>
ATM	<i>Ataxia Telangiectasia Mutated</i>
a-EJ	<i>Alternative-End Joining</i>
BAC	<i>Bacterial Artificial Chromosome</i>
BAM	<i>Binary Alignment Map</i>
BARD1	<i>BRCA1 Associated Ring Domain 1</i>
BER	<i>Base Excision Repair</i>
BRCA1	<i>Breast Cancer susceptibility protein-1</i>
BRCA2	<i>Breast Cancer susceptibility protein-2</i>
BRCT	<i>BRCA1 Carboxy-Terminal domain</i>
CBP	<i>Creb Binding Protein</i>
CC	<i>Coiled-Coil</i>
CDC2	<i>Cell Division Cycle protein 2</i>
CDK	<i>Cyclin-Dependent Kinases</i>
CMT	<i>Cancer Médullaire de la Thyroïde</i>
CNV	<i>Copy Number Variations</i>
CRT	<i>Cyclic Reversible Termination</i>
DBD	<i>DNA Binding Domain</i>
DRAM	<i>Damage-Regulated Autophagy Modulator</i>
DSB	<i>Double-Strand Break</i>
dTTP	<i>DésoxyThymidine TriPhosphate</i>
dUTP	<i>DésoxyUridine TriPhosphate</i>
EASE	<i>Expression Analysis Systematic Explorer</i>
EBV	<i>Epstein – Barr Virus</i>
EDTA	<i>Acide Ethylène-Diamine-Tétra-Acétique</i>
EGFP	<i>Enhanced Green Fluorescent Protein</i>
EMT	<i>Epithelial-Mesenchymal Transition</i>
EMSY	<i>BRCA2 interacting transcriptional repressor</i>
emPCR	<i>PCR En Emulsion</i>
ENCODE	<i>Encyclopedia Of DNA Elements</i>
FA	<i>Anémie de Fanconi</i>
FASAY	<i>Functional Analysis of Separated Alleles in Yeast</i>
GFP	<i>Green Fluorescent Protein</i>
GnomAD	<i>Genome Aggregation Database</i>
GOF	<i>Gain Of Function</i>
GSP	<i>Genome Sequencing Program</i>
GTF	<i>Gene Transfer Format</i>
GTP	<i>Guanosine TriPhosphate</i>
HBOC	<i>Hereditary Breast and Ovarian Cancer</i>

HDAC1	<i>Histone DeAcetylase 1</i>
HDR	<i>Homology-Directed Repair</i>
HNF3	<i>Hepatocyte Nuclear Factor 3</i>
HNPCC	<i>Hereditary NonPolyposis Colorectal Cancer</i>
HR	<i>Homologous Recombination</i>
HTS	<i>High-Throughput Sequencing</i>
IARC	<i>International Agency for Research on Cancer</i>
ICL	<i>Interstrand CrossLinks</i>
IGV	<i>Integrative Genomic Viewer</i>
IMS	<i>Instabilite MicroSatellitaire</i>
INCa	<i>Institut National du Cancer</i>
iPSC	<i>Induced Pluripotent Stem Cells</i>
IRIF	<i>Irradiation Induced Foci</i>
IRM	<i>Imagerie par Résonance Magnétique</i>
ISP	<i>Ion Sphere Particule</i>
LFS	<i>Li-Fraumeni Syndrome</i>
LOF	<i>Loss Of Function</i>
LOH	<i>Loss Of Heterozygosity</i>
MLPA	<i>Multiplex Ligation-dependent Probe Amplification</i>
MMR	<i>MisMatch Repair</i>
MRN	<i>Mre11, Rad50, Nbs1</i>
mRNP	<i>RiboNucléoProtéines Messagères</i>
NaCl	<i>Chlorure de Sodium</i>
NER	<i>Nucleotide Excision Repair</i>
NES	<i>Nuclear Export Signal</i>
NF-Y	<i>Nuclear transcription Factor Y</i>
NGS	<i>Next Generation Sequencing</i>
NLS	<i>Nuclear Localization Signal</i>
NHEJ	<i>Non-Homologous End Joining</i>
NMD	<i>Non-sense Mediated Decay</i>
OB	<i>Oligonucleotide Binding</i>
ONT	<i>Oxford Nanopore Technologies</i>
OPLS-DA	<i>Orthogonal Partial Least Square Discriminant Analysis</i>
PALB2	<i>Partner And Localizer of BRCA2</i>
PARP	<i>PolyAdenosine diphosphate-Ribose Polymerase</i>
Pb	<i>Paire de Bases</i>
PBMC	<i>Peripheral Blood Mononuclear Cell</i>
PCNA	<i>Proliferating Cell Nuclear Antigen</i>
PCR	<i>Polymerase Chain Reaction</i>
PIC	<i>Pre-Initiation complex</i>
PML	<i>ProMyelocytic Leukemia protein</i>
PPi	<i>PyroPhosphate Inorganique</i>
QMPSF	<i>Quantitative Multiplex PCR of Short Fluorescent fragments</i>
Rb	<i>RetinoBlastoma protein</i>
RE	<i>Response Element</i>
RGC	<i>Ribosome Gene Cluster</i>
RIN	<i>RNA Integrity Number</i>
RNA _{si}	<i>RNA Small Interfering</i>
ROS	<i>Reactive Oxygen Species</i>
RPA	<i>Replication Protein A</i>

RT	<i>Reverse Transcription</i>
RT-MLPA	<i>Reverse Transcription Multiplex Ligation-dependant Probe Amplification</i>
RT-QMPSF	<i>Reverse Transcription – Quantitative Multiplex PCR of Short Fluorescent fragments</i>
SAC	<i>Spindle Assembly Checkpoint</i>
SBS	<i>Sequencing By Synthesis</i>
SCD	<i>Serine Cluster Domain</i>
SH3	<i>Src Homology 3</i>
SMRT	<i>Single Molecule Real Time</i>
SNA	<i>Single Nucleotide Addition</i>
SNV	<i>Single Nucleotide Variant</i>
SRA	<i>Sequence Read Archive</i>
SSA	<i>Single Strand Annealing</i>
STAR	<i>Spliced Transcripts Alignment to a Reference</i>
SVR	<i>Support Vector Regression</i>
TA	<i>Transcription Activation</i>
TAD	<i>Trans-Activating Domain</i>
TET	<i>Tetramerization Domain</i>
TERT	<i>Telomerase Reverse Transcriptase</i>
TP53	<i>Tumor Protein 53</i>
TSS	<i>Transcription Start Site</i>
UA	<i>Unités Arbitraires</i>
UCSC	<i>University of California Santa Cruz</i>
UDG	<i>Uracil DNA Glycosylase</i>
UHPLC-IM-MS	<i>Ultra-High Performance Liquid Chromatography – Ionic Mobility – Mass Spectrometry</i>
UMI	<i>Unique Molecular Identifier</i>
UTR	<i>UnTranslated Regions</i>
VCF	<i>Variant Call Format</i>
VSI	<i>Variation de Signification biologique Inconnue</i>
WES	<i>Whole-Exome Sequencing</i>
WGS	<i>Whole-Genome Sequencing</i>
XIST	<i>X Inactive Specific Transcript</i>
XPB	<i>Xeroderma Pigmentosum groupe B</i>
XPC	<i>Xeroderma Pigmentosum group C</i>
XPD	<i>Xeroderma Pigmentosum groupe D</i>
ZMW	<i>Zero-Mode Waveguides</i>
2D	<i>Two Dimensions</i>
3D	<i>Three Dimensions</i>

Introduction

Chapitre I

Les prédispositions héréditaires au cancer

I - Le cancer, une maladie génétique

Notre génome subit constamment de multiples lésions. Si des facteurs environnementaux tels que les mutagènes, les rayons ionisants ou ultra-violetts peuvent être à l'origine de ces altérations génétiques, les processus physiologiques tels que la réplication ou le métabolisme sont également à l'origine de nombreuses altérations de notre ADN (acide désoxyribonucléique). Pour permettre le maintien de l'intégrité du génome et une transmission fidèle de l'information génétique aux cellules filles de l'organisme ainsi qu'à la génération précédente, ces altérations génétiques sont prises en charge par les systèmes de réparation de l'ADN. En effet, suite à l'apparition de lésions sur le génome, les points de contrôle du cycle cellulaire sont activés permettant ainsi l'arrêt du cycle cellulaire et la réparation des lésions par les systèmes spécifiques des altérations détectées. Si les lésions sont trop nombreuses et que les systèmes de réparation sont dépassés, la cellule peut enclencher un processus de mort cellulaire programmée pour éviter de transmettre les mutations aux cellules filles. Au fil des divisions cellulaires et du vieillissement de l'individu, ces lésions somatiques peuvent échapper aux systèmes de réparation de l'ADN et ainsi s'accumuler dans notre génome. Cette accumulation de mutations lorsqu'elle survient dans des gènes clés peut provoquer une instabilité génétique et un dysfonctionnement cellulaire conférant un avantage sélectif à la cellule pouvant être à l'origine du développement tumoral.

Ces gènes clés pour le développement tumoral sont actuellement classés en deux grandes catégories : les **proto-oncogènes** et les **gènes suppresseurs de tumeurs** regroupant les « gate-keepers » et les « care-takers » (Vogelstein & Kinzler, 2004).

I. Les proto-oncogènes

La première classe de gènes clés pour la transformation tumorale englobe les **proto-oncogènes**. Les proto-oncogènes sont des gènes **activant la cascade de transduction du signal**. Des mutations somatiques ou constitutionnelles peuvent être à l'origine d'une activation inappropriée de ces gènes que l'on nomme alors **oncogènes**. Les proto-oncogènes ont été découverts dans les années 1970, par des recherches menées sur les rétrovirus cancérogènes qui possèdent le gène v-Src engendrant des cancers chez le poulet et son homologue humain le gène c-src (Stehelin et al., 1976). Cette découverte a valu le prix Nobel de Médecine à John Michael Bishop et à Harold Varmus en 1989.

Les oncogènes peuvent être classés en cinq classes : les gènes codant pour des facteurs de croissance (ex : *FGF3*, *Fibroblast Growth Factor 3*), les gènes codant pour les récepteurs aux facteurs de croissance (ex : le récepteur de l'EGF, (*Epidermal Growth Factor*) du PDGF

(*Platelet-Derived Growth Factor*) ou du VEGF (*Vascular Endothelial Growth Factor*), les gènes codant pour des protéines impliquées dans la transduction du signal (ex : la Sérine/thréonine kinase Raf1, *Rapidly Accelerated Fibrosarcoma 1* ; la GTPase, Guanosine TriPhosphate ; Kras, *Kirsten rat sarcoma viral oncogene homolog*), des gènes codant pour des facteurs de transcription (ex : c-Myc, *Cellular MYeloCytomatosis* ; c-Fos ; c-Jun) et des gènes codant pour des régulateurs d'apoptose (ex : BCL2, B-cell CLL Lymphoma 2). Pour contribuer à la tumorigenèse, l'activité des proto-oncogènes doit être modifiée par des mutations activatrices permettant une activation constitutionnelle inappropriée de ces gènes impliqués dans la croissance cellulaire. Cette augmentation d'activité peut également résulter d'une amplification génique ou d'une dérégulation de l'expression de ces gènes par des mécanismes de translocations ou de réarrangements chromosomiques. Comme exemple, l'activation dérégulée des GTPases telles que Ras (Rous avian sarcoma) qui jouent un rôle important dans la transmission vers le noyau de signaux extracellulaires provenant de récepteurs membranaires, notamment de l'EGFR (Epidermal Growth Factor Receptor), provoque l'activation de la voie MAPK (Mitogen-Activated Protein Kinase) et la signalisation incontrôlée, aboutissant aux phénomènes de prolifération, invasion, migration, etc. L'altération d'un seul allèle de cette classe de gène est suffisante pour conférer un avantage à la cellule.

II. Les gènes suppresseurs de tumeurs

La deuxième classe regroupe **les gènes suppresseurs de tumeurs**. Ces gènes sont responsables **du maintien de l'intégrité du génome et du contrôle de la croissance et de la prolifération cellulaire**. L'activité de ces gènes est altérée dans les cancers par des mutations inactivatrices, perte de fonction (LOF, *Loss Of Function mutations*). Selon le modèle de Knudson (Knudson, 1971), les deux allèles du gène suppresseur de tumeur doivent être mutés pour contribuer à la progression tumorale. Cette théorie des « deux hits » ou « deux événements » publiée en 1971 par Alfred G. Knudson a été décrite à partir d'une analyse épidémiologique réalisée par ce généticien américain sur des rétinoblastomes (tumeur pédiatrique de l'œil) sporadiques et héréditaires liées à des mutations du gène *RBI* (Retinoblastoma Transcriptional Corepressor 1). Dans les formes héréditaires de rétinoblastome, une mutation constitutionnelle est déjà présente dans ces cellules rétinienne et il suffit d'un seul événement somatique supplémentaire touchant l'autre allèle pour développer la tumeur, à un âge précoce. Dans les cas sporadiques, le développement de la tumeur résulte de l'inactivation somatique des deux allèles au cours du développement de

l'œil et sont donc associés à un âge de survenue plus tardif. Pour les gènes suppresseurs de tumeurs, le premier allèle est inactivé par des mutations perte de fonction et l'allèle restant est souvent perdu lors d'un remaniement génétique qui conduit à la délétion d'une partie plus ou moins importante du chromosome (perte d'hétérozygotie ou LOH, Loss Of Heterozygosity).

Les gènes suppresseurs de tumeurs peuvent être séparés en deux catégories principales : les *gatekeepers* et les *caretakers*.

Les gènes *gatekeepers* agissent sur les voies de **contrôle de la division cellulaire**, en induisant l'arrêt du cycle cellulaire ou différents types de mort cellulaire (apoptose, autophagie, ferroptose). Parmi ces gènes, on compte par exemple le gène suppresseur de tumeurs *PTEN* (*Phosphatase and TENsin homolog*) qui inhibe l'oncogène *Akt* (*serine/threonine kinase*) dans la voie de signalisation de l'EGF (*Epidermal Growth Factor*), le gène *BIM* (*Proapoptotic BH3⁻only*), faisant partie de la famille des Bcl-2 (*B-cell lymphoma 2*), qui favorise l'apoptose ainsi que le gène suppresseur de tumeurs *TP53* (*Tumor Protein 53*) qui tient un rôle clef dans la régulation du cycle cellulaire et de la mort cellulaire.

Les *caretakers* ne régulent pas directement le contrôle du cycle cellulaire, mais agissent pour maintenir la stabilité génomique en jouant un rôle dans la **réparation des lésions de l'ADN**. En effet, ces gènes sont impliqués dans les grandes voies de réparation de l'ADN afin d'éliminer physiquement les lésions. Parmi ces grandes voies de réparation on distingue les systèmes suivants :

- La réparation par excision de base (BER, *Base Excision Repair*) est le système impliqué dans la réparation des bases oxydées, alkylées, désaminées mais aussi dans la réparation des cassures simple brin de l'ADN (Dianov & Hübscher, 2013; Chatterjee & Walker, 2017).
- La réparation par excision de nucléotides (NER, *Nucleotide Excision Repair*) est la voie chargée d'éliminer les lésions volumineuses telles que les dimères de cyclobutaniques pyrimidines résultant de l'exposition aux rayonnements ultraviolets, tels que les adduits de benzo[a]pyrène ou encore ceux causées par les agents chimio thérapeutiques (Wakasugi *et al.*, 2002; Chatterjee & Walker, 2017).
- La réparation des mésappariements de l'ADN (MMR, *Mismatch Repair*) qui est une voie de réparation post-répllicative en charge de corriger les erreurs de répllication (Jiricny, 2013; Crouse, 2016, p.).
- La réparation par les protéines de la voie de l'anémie de Fanconi pontages intra-brins ou inter-brins causés par des agents réticulants tels que les composés du platine et les agents alkylants (Michl *et al.*, 2016).

- La réparation des cassures double brin (DSB, *Double-Strand Break*) de l'ADN par les deux principales voies qui suppriment la majorité des DSB : la recombinaison non homologe (NHEJ, *Non-Homologous End Joining*), et la recombinaison homologe (HR, *Homologous Recombination*). Un autre mécanisme alternatif de jonction d'extrémités (a-EJ, *Alternative-End Joining*) apporte aussi une contribution mais mineure à cet égard (Chatterjee & Walker, 2017; Sung, 2018) (**Figure 1**).

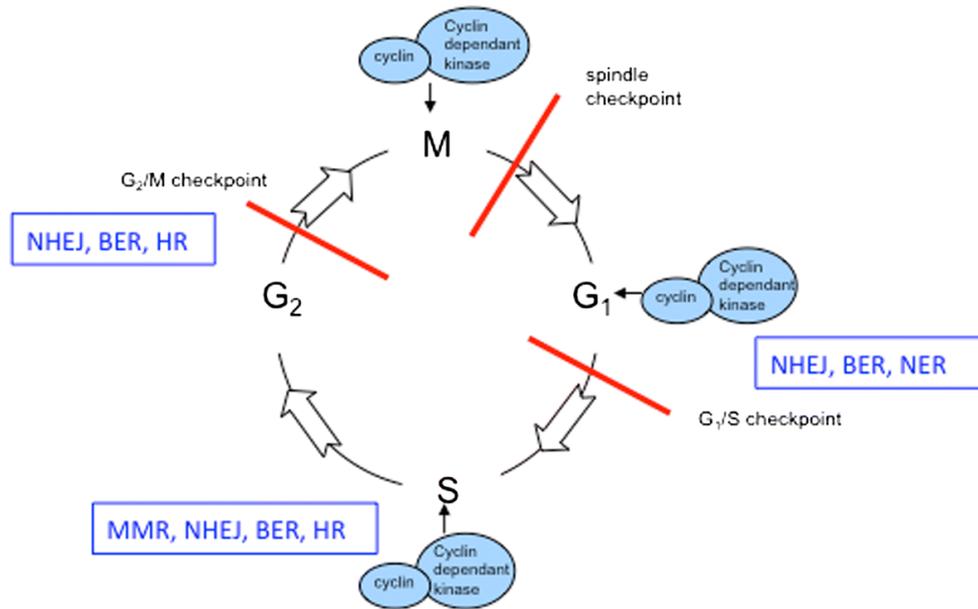


Figure 1 : Le moment d'intervention des différents systèmes de réparation de l'ADN au cours du cycle cellulaire (D'après Gee et al., 2018)

Le système de réparation par excision de base (BER) ainsi que le système de réparation par la recombinaison non homologe (NHEJ) réparent l'ADN tout au long du cycle cellulaire. La réparation par excision de nucléotides (NER) est effectuée au cours de la phase G₁, la réparation des mésappariements (MMR) au cours de la phase S et la recombinaison homologe (HR) répare les cassures double brin de l'ADN pendant les phases S et G₂.

L'accumulation d'altérations génomiques séquentielles au niveau de ces classes de gènes aboutit à la conversion d'une cellule normale en une cellule cancéreuse. Ce processus de transformation du cancer s'accompagne de caractéristiques émergentes parmi lesquelles le maintien de la signalisation proliférative, l'autosuffisance et l'insensibilité aux signaux de croissance, la résistance à l'apoptose, l'activation de la sénescence, l'apparition d'un potentiel répliatif illimité, la capacité à induire l'angiogenèse et l'invasion tissulaire, et l'activation des processus métastatiques (Hanahan & Weinberg, 2000; Fouad & Aanei, 2017) (**Figure 2**).

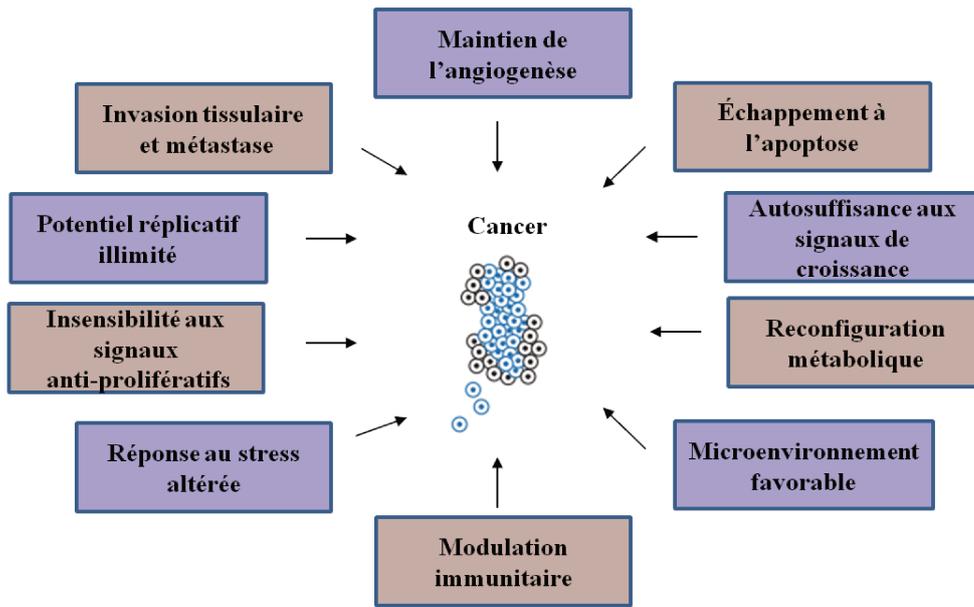


Figure 2 : Capacités acquises par les cellules cancéreuses

Les études de séquençage à haut-débit des tumeurs ont révélé l'existence d'une centaine d'altérations somatiques dans une cellule tumorale. Seule une minorité de ces mutations sont appelées *driver mutations*. Une tumeur typique contient deux à huit mutations *driver* jouant un rôle causal dans la transformation tumorale. Le reste sont des mutations dites passagères ou *passenger mutations* ne conférant pas d'avantage sélectif à la tumeur mais qui reflètent l'instabilité génétique de la cellule cancéreuse. L'ordre d'apparition des mutations *driver* semble être essentiel au développement tumoral. Cette accumulation progressive d'anomalies génétiques détectables dans une population de cellules va provoquer une hétérogénéité génétique au sein d'une même tumeur, avec des zones tumorales ne présentant pas toutes les mêmes altérations. Ainsi, en réalisant plusieurs analyses génétiques sur plusieurs zones de la tumeur, il est possible de réaliser une véritable phylogénie mutationnelle en déterminant l'ordre d'apparition des mutations au cours du développement de la tumeur (Hong et al., 2015; Schwartz & Schäffer, 2017) (**Figure 3**).

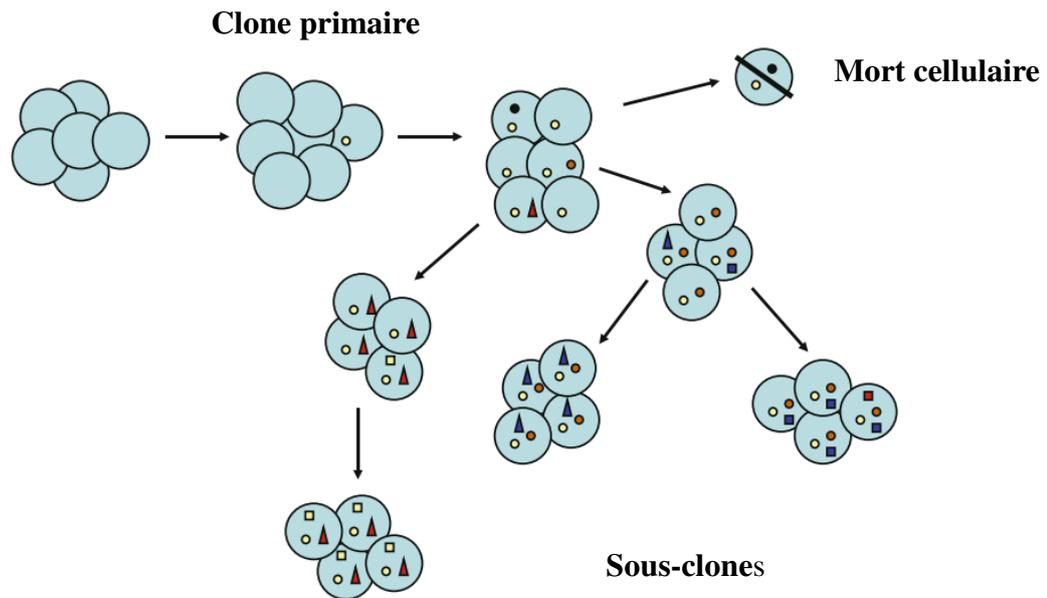


Figure 3 : Phylogénie génétique des tumeurs (Adaptée de Braekeleer et al., 2015)

L'évènement mutationnel causal produit dans une cellule est à l'origine de la formation de la première cellule tumorale qui va s'expandre pour former le premier clone. Des mutations somatiques supplémentaires peuvent survenir dans certaines cellules du clone, conduisant à des sous-clones dans lesquels certaines cellules peuvent également héberger de nouvelles mutations uniques. Ce processus de ramification de l'évolution de la lignée cancéreuse génère une généalogie hiérarchique qui peut être analysée à l'aide des outils phylogénétiques de la biologie de l'évolution.

II - Formes mendéliennes de cancer

Sur l'ensemble des cancers, 90 % correspondent à des cancers sporadiques, causés par l'accumulation séquentielle de mutations somatiques au cours du temps. Si la très grande majorité des cas de cancer correspondent à des cas sporadiques liés à l'accumulation au cours du vieillissement de mutations somatiques, 5 à 10 % des cancers correspondent à de véritables maladies mendéliennes, liées à des mutations constitutionnelles dans les gènes de prédisposition au cancer. Ces cas représentent 1,4 million de nouveaux cas / an à l'échelle mondiale (Kulkarni & Carley, 2016).

Ces prédispositions génétiques au cancer sont liées à l'activation constitutionnelle d'oncogènes ou à l'inactivation du gène suppresseur de tumeur par une mutation constitutionnelle suivie d'une inactivation du deuxième allèle par un évènement somatique. Ces altérations vont créer un environnement permissif favorisant l'accumulation d'autres altérations somatiques. Ces formes mendéliennes se caractérisent donc par un risque accru de cancer chez les individus porteurs de mutations délétères par rapport à la population générale ainsi que par la précocité d'apparition des tumeurs.

Pour effectuer le diagnostic de ces pathologies, des généticiens se basent sur des caractéristiques majeures de ces prédispositions génétiques.

► **L'histoire personnelle du patient** : diagnostic à un âge inhabituel (précoce, généralement avant l'âge de 51 ans), existence de cancers primitifs multiples chez le patient, localisation des tumeurs, association avec des malformations.

► **Les antécédents familiaux** : atteinte de deux ou plusieurs individus de la même famille par le même type de cancer et détection du cancer sur plusieurs générations.

L'identification des cas évocateurs d'une prédisposition génétique est parfois complexifiée par plusieurs facteurs : les familles sont souvent de petite taille, la pénétrance des mutations est souvent incomplète et il existe également une variabilité phénotypique et des risques de cancer différents entre les sexes.

Il est donc nécessaire pour les cliniciens d'établir pour chaque type de prédisposition les critères nosologiques permettant d'identifier les individus évocateurs qui pourront alors bénéficier d'une analyse génétique. Ces critères ont pour but d'identifier le maximum d'individus porteurs de mutations délétères tout en limitant de façon rationnelle et pragmatique le nombre d'individus qui seront soumis à un test génétique. Ces critères nosologiques évoluent avec les connaissances acquises sur la maladie génétique mais

également avec les changements de pratiques cliniques. Comme exemple de critères, on peut citer les recommandations internationales de bonnes pratiques (guidelines) standardisées, qui ont été établies pour faciliter le diagnostic du syndrome de Lynch. Ces critères ont évolué au cours du temps (critères d'Amsterdam I, d'Amsterdam II, de Bethesda) et la dernière version correspond aux critères II de Bethesda modifiés (**Tableau 1**).

Table 1. Amsterdam criteria and Bethesda guidelines for a positive Lynch syndrome diagnosis

Amsterdam I criteria

1. Three or more relatives with histologically confirmed CRC, with one being a first-degree relative of the other two provided that FAP has been excluded
2. Two or more generations with CRC
3. One or more CRC cases diagnosed before the age of 50 years

Amsterdam II criteria

1. Three or more relatives with histologically confirmed HNPCC-associated cancers (CRC, cancer of the endometrium, small bowel, renal pelvis, or ureter) with one being a first-degree relative of the other two provided that FAP has been excluded
2. Two or more generations involved
3. One or more cancers diagnosed before the age of 50 years

Revised Bethesda guidelines

1. Diagnosed with colorectal cancer before the age of 50 years
2. Synchronous or metachronous colorectal or other HNPCC-related tumors (including stomach, ureter, renal pelvis, and biliary tract cancers, glioblastoma, sebaceous gland adenoma, keratocanthoma, and carcinoma of the small bowel) regardless of age
3. Colorectal cancer with a high-MSI morphology that was diagnosed before the age of 60 years
4. Colorectal cancer with one or more first-degree relatives with colorectal cancer or other HNPCC-related tumors. One of the cancers must have been diagnosed before the age of 50 years (this includes adenoma, which must have been diagnosed before the age of 40 years)
5. Colorectal cancer with two or more relatives with colorectal cancer or other HNPCC-related tumors regardless of age

CRC, colorectal cancer; FAP, familial adenomatous polyposis patients; HNPCC, hereditary nonpolyposis colorectal cancer; MSI, microsatellite instability.

Tableau 1 : Description des critères d'Amsterdam I et II et des revised Bethesda guidelines (D'après Pradere et al., 2017)

Aujourd'hui, plus de 80 gènes de prédisposition au cancer ont été identifiés (Institut National du Cancer, INCA) (**Tableau 2**). Cette identification a été rendue possible par des approches de liaisons génétiques menées dans des grandes familles, par des approches de gènes candidats et plus récemment par des stratégies d'identification basées sur le séquençage à haut-débit (NGS, Next Generation Sequencing) de l'exome ou du génome. De façon intéressante, une fraction importante des gènes de prédisposition au cancer sont également mutés dans les cancers sporadiques (Rahman, 2014).

L'identification des bases moléculaires de ces formes héréditaires de cancer a grandement contribué à une meilleure compréhension de ces pathologies mais également des formes sporadiques de cancer. Elle permet également d'ouvrir la voie vers le développement de thérapies spécifiques. Parmi les syndromes de prédisposition héréditaire au cancer, le syndrome de Lynch ou cancer colorectal héréditaire non polyposique (HNPCC, *Hereditary NonPolyposis Colorectal Cancer*) et le cancer héréditaire du sein et de l'ovaire sont les deux formes les plus fréquentes.

PRÉDISPOSITIONS GÉNÉTIQUES	PRINCIPAUX GÈNES ASSOCIÉS
SYNDROME SEINS-OVAIRES	BRCA1, BRCA2, PALB2, RAD51
SYNDROME DE LYNCH	MLH1, MSH2, MSH6, PMS2, EPCAM
ADÉNOMES HYPOPHYSAIRES FAMILIAUX	AIP
ATAXIE-TÉLANGIECTASIE	ATM, MRE11A
CANCER GASTRIQUE DIFFUS FAMILIAL	CDH1
CARCINOME PAPILLAIRE RÉNAL HÉRÉDITAIRE	FH, MET
HYPERPARATHYROIDISME	CDC73, CASR
MALADIE DE COWDEN	PTEN, PIK3CA, AKT1
MALADIE DE FANCONI	FANC
MALADIE DE VON HIPPEL-LINDAU	VHL
MÉLANOME MALIN FAMILIAL	CDKN2A, MITF, BAP1, POT1, CDK4
NÉOPLASIES ENDOCRINIENNES	MEN1, RET, CDKN1B
NEUROFIBROMATOSES	NF1, NF2, LZTR1, SMARCB1, SPRED1, SMARCE1
PHÉOCHROMOCYTOME-PARAGANGLIOME HÉRÉDITAIRE	SDH, TMEM127, MAX, EPAS1
POLYPOSES ADÉNOMATEUSES FAMILIALES	APC, MUTYH, POLE, POLD1, NTHL1
RÉTINOBLASTOME	RB1
SYNDROME DE BIRT-HOGG-DUBÉ	FLCN
SYNDROME DE BLOOM	BLM
SYNDROME DE CARNEY	PRKAR1A, ARMC5
SYNDROME DE GORLIN	PTCH1, PTCH2, SUFU
SYNDROME DE LI-FRAUMENI	TP53, CHEK2
SYNDROME DE NIJMEGEN	NBN
SYNDROME DE PEUTZ-JEGHERS	STK11
SYNDROME DE POLYPOSE JUVÉNILE	BMPR1A, SMAD4
SYNDROME DE WERNER	WRN
XERODERMA PIGMENTOSUM	XP

Tableau 2 : Les principales prédispositions génétiques au cancer et les principaux gènes associés à ces prédispositions (Données issues de l'INCA, Institut National du Cancer, 2018, <https://www.e-cancer.fr/>)

III- Diagnostic des formes héréditaires de cancer

Les syndromes de prédisposition génétique au cancer ont fourni des modèles pour approfondir notre compréhension du développement du cancer et constituent un important problème de santé en soi. Il est crucial de pouvoir identifier de façon précoce les familles à risque élevé de cancer pour pouvoir leur proposer un suivi médical adapté. Ce suivi médical consiste en un dépistage précoce des tumeurs, la possibilité d'interventions chirurgicales en temps adéquat et à l'accès à des thérapies ciblées appropriées. Il existe donc une étroite collaboration entre les oncologues, les radiologues et les généticiens cliniques en charge d'identifier ces familles à risque. Pour coordonner ces efforts, des groupes de travail nationaux ou internationaux spécialistes d'une prédisposition, travaillent à élaborer et à mettre à jour les critères nosologiques permettant d'identifier le plus d'individus à risque tout en limitant le nombre de tests génétiques réalisés.

En premier lieu, lorsqu'un clinicien suspecte une prédisposition génétique, il est amené à adresser le patient en **consultation d'oncogénétique** afin de rencontrer un médecin spécialiste de la génétique des cancers. Le généticien ou le conseiller en génétique va **recueillir les données familiales** et tenter d'obtenir des informations ou des confirmations histologiques sur la tumeur du patient. Si l'hypothèse d'une prédisposition génétique est retenue, le généticien apportera au patient une explication détaillée des risques de cancer et des directives de prise en charge. Les patients seront également informés des implications familiales potentielles des résultats du bilan diagnostique. Après signature d'un **consentement éclairé**, une **analyse génétique** est alors envisagée. Un premier prélèvement sanguin est alors effectué accompagné ou non d'un second prélèvement correspondant à un frottis jugal. Ce prélèvement est adressé au laboratoire de génétique moléculaire où sera réalisé le test génétique. En cas de résultat positif, une confirmation sera réalisée sur un deuxième prélèvement indépendant. L'identification de la mutation constitutionnelle à l'origine de la prédisposition permet dans un premier temps de **poser le diagnostic** chez le patient et de réaliser un **conseil génétique** pour la famille.

1- Impact sur les apparentés

Lorsque la ou les mutations causales de la prédisposition sont mises en évidence dans la famille à partir d'un sujet atteint, un test génétique est proposé aux apparentés à risque pour identifier les apparentés porteurs qui nécessitent une surveillance médicale adaptée et les

apparentés non porteurs qui ont un risque de cancer équivalent à celui de la population générale et chez qui cette surveillance est inappropriée. Les membres de la famille porteurs de la mutation délétère mais encore asymptomatiques au moment du test génétique peuvent être identifiés et bénéficier d'un conseil génétique et d'un suivi médical visant à prévenir et à dépister précocement l'apparition des tumeurs.

2- Impact sur le patient

Le patient porteur de la mutation peut se voir proposer un plan visant à réduire le risque de cancer ou à diagnostiquer le cancer à un stade précoce plus facile à gérer. Durant cette phase d'intervention, les principales options de gestion du risque de cancer chez ces patients comprennent les éléments listés ci-dessous.

a- Un dépistage du cancer et une surveillance accrue et appropriée

L'identification d'un syndrome héréditaire de prédisposition au cancer peut aider les patients à accéder à des données probantes de programmes de **dépistage**. Ces programmes ont pour but de faciliter la détection précoce des tumeurs associées et prévenir leur apparition pour favoriser les chances de guérison et ainsi diminuer la mortalité.

Chez les femmes porteuses d'une mutation du gène *BRCA1/2* (Breast Cancer 1/2) et ainsi prédisposées au cancer du sein, le dépistage consiste à une palpation des seins et des creux axillaires tous les 6 mois et une IRM mammaire annuelle à partir de 30 ans ou 5 ans avant le cancer du sein le plus précoce dans la famille s'il existe un cas avant 30 ans.

Malgré la réalisation de cette IRM à un rythme annuel, des cancers de seins ont été révélés pendant un intervalle d'une année entre deux IRM effectuées. D'où la proposition aux femmes porteuses d'une mutation des gènes *BRCA* d'effectuer une mastectomie prophylactique.

Pour les porteurs d'une mutation dans l'un des gènes *MMR* présentant un risque élevé de développer le syndrome de Lynch, le suivi doit débuter dès l'âge de 20 ans. La surveillance doit être effectuée par endoscopie colorectale complète avec chromoscopie par indigo-carmin, réalisée tous les 2 ans, en insistant sur la bonne préparation colique.

b- Une chirurgie prophylactique

Après l'identification d'une mutation délétère, une **chirurgie prophylactique** pourra être envisagée en fonction des pathologies et des risques de cancer associés. Cette chirurgie constitue un temps particulièrement fort du parcours médical de ces patients et de leur entourage. Plusieurs approches ont été recommandées :

- Pour le cancer du sein, malgré son caractère mutilant, la mastectomie bilatérale prophylactique reste la mesure la plus efficace de prévention du risque de cancer du sein chez une femme indemne porteuse d'une mutation *BRCA1/2*.
- Pour le cancer gastrique, une gastrectomie totale prophylactique est recommandée chez un sujet porteur d'une mutation du gène *CDH1 (Cadherin 1)*.
- Pour le cancer médullaire de la thyroïde (CMT), il est recommandé de réaliser une thyroïdectomie associée à un curage ganglionnaire central récurrentiel et latérocervical bilatéral systématique chez les porteurs d'une mutation constitutionnelle du gène *RET*.

La prise en charge des patients pour lesquels se pose la question d'une intervention chirurgicale prophylactique implique fortement l'équipe médicale qui doit accompagner le patient dans la formulation de sa demande et dans sa décision finale. Les avantages et les risques relatifs de chacun de ces choix sont expliqués en détail au patient. Même si le bénéfice médical de cette prise en charge pré-symptomatique est le plus souvent majeur, il est important de proposer un **accompagnement psychologique** afin que ce bénéfice médical ne soit pas contrebalancé par un impact psychologique négatif.

c- Une adaptation des stratégies thérapeutiques

Différentes combinaisons de schémas thérapeutiques susceptibles d'améliorer considérablement le pronostic et les résultats de la maladie sont en cours d'évaluation. A titre d'exemple, dans les cancers colorectaux à instabilité microsatellitaire (MSI), les patients présentant des mutations décalant le cadre de lecture de certains gènes suite à un défaut de fonctionnement du système MMR, pourraient être la cible de stratégies d'immunothérapie cellulaire. Grâce à des essais *in vitro* utilisant des cellules présentatrices d'antigène artificielles codant les néoantigènes dérivant de ces mutations, notre équipe a effectué une stimulation et une amplification de lymphocytes T CD8+ cytotoxiques fonctionnels et spécifiques de ces néoantigènes dérivant de mutations qui entraînent un décalage du cadre de lecture. Ces résultats soulignent l'importance d'identifier les mutations pouvant générer des

néoantigènes immunogènes pour le développement de nouvelles approches d'immunothérapie cellulaire adoptive (Maby *et al.*, 2015).

Une autre approche très prometteuse et ayant déjà reçu une autorisation de mise sur le marché pour certains types de cancer exploite les faiblesses des cellules tumorales des patientes porteuses d'une altération constitutionnelle des gènes *BRCA1* ou *BRCA2*. En effet, ces cellules sont déficientes pour la voie de recombinaison homologue. Pour optimiser la destruction des cellules tumorales, le concept de létalité synthétique est utilisé et des inhibiteurs de la polyadénosine diphosphate-ribose polymérase (PARP1) sont utilisés en conjonction avec la chimiothérapie conventionnelle. L'enzyme nucléaire PARP1 joue un rôle important dans la réparation des cassures simple brin de l'ADN via le système BER. Les inhibiteurs de PARP vont bloquer cette réparation et les cassures monocaténaire non réparées seront converties en cassures double brin au cours de la réplication de l'ADN. Ces cassures pourront être prises en charge par un système encore opérationnel dans les cellules normales des patientes mais ne pourront pas être réparées de façon efficace et fiable dans les cellules tumorales déficientes en *BRCA* (**Figure 4**). L'instabilité génomique générée rend les cellules tumorales beaucoup plus sensibles aux lésions introduites par la chimiothérapie (Bryant *et al.*, 2005; Farmer *et al.*, 2005; Lord & Ashworth, 2013). Plusieurs essais thérapeutiques ont démontré l'efficacité des inhibiteurs de PARP dans le traitement des cancers du sein et de l'ovaire. L'olaparib est le premier inhibiteur de PARP à avoir obtenu une autorisation de mise sur le marché européen (AMM) en décembre 2014. Il est destiné aux patientes atteintes d'un cancer de l'ovaire et porteuses d'une mutation *BRCA* après au moins deux lignes de chimiothérapies à base de platine (Kim *et al.*, 2015a; Lorusso *et al.*, 2018). Des essais cliniques sont en cours afin d'autoriser son application chez des patients atteints de cancer du sein (Robson *et al.*, 2017; Tong *et al.*, 2018).

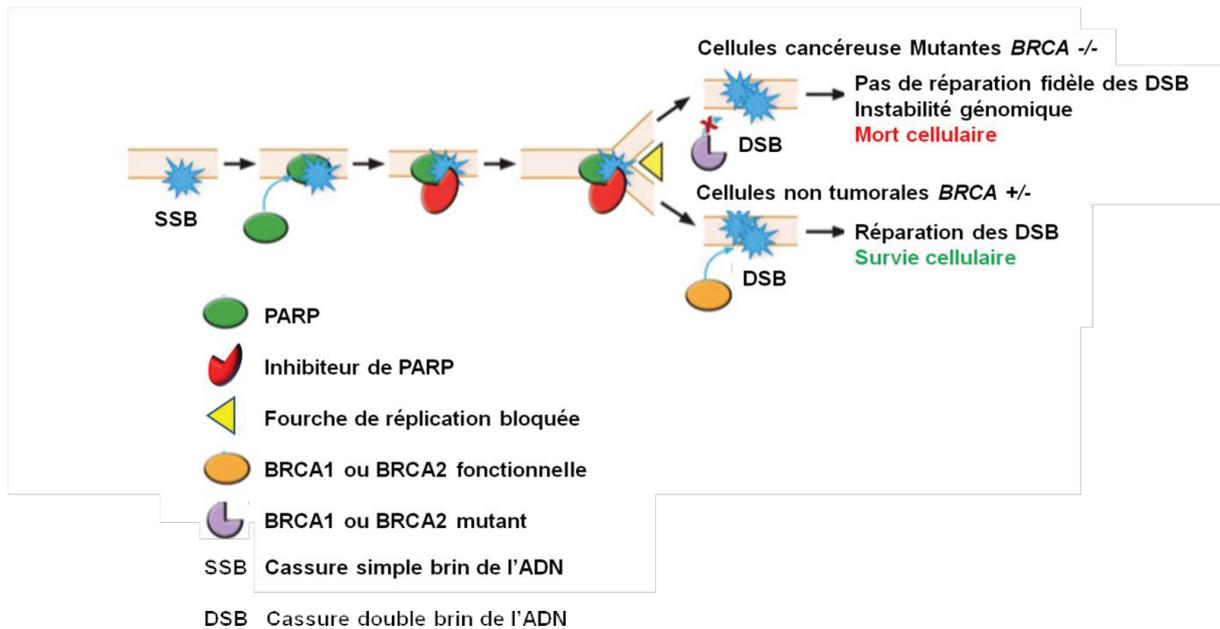


Figure 4 : Schéma présentant le principe de fonctionnement des inhibiteurs de PARP (Figure adaptée de Turk & Wisinski, 2018)

Les protéines PARP sont principalement impliquées dans la réparation des cassures simple brin (SSB) de l'ADN. L'ajout des inhibiteurs de PARP dans ces cellules en inhibant la réparation des cassures simple brin, va générer des cassures double brin (DSB) et un blocage de la fourche de réplication. Ces lésions sont réparées par la voie de recombinaison homologue (HR) impliquant les protéines BRCA. Les cellules cancéreuses mutantes pour l'un des deux gènes *BRCA* présentent une HR altérée, seront sensibilisées à l'inhibition de l'activité enzymatique de PARP vu qu'elles ne seront pas capables de réparer les DSB de façon fidèle. Cela entrainera ainsi une instabilité chromosomique et par la suite la mort des cellules cancéreuses. D'autre part, dans les cellules non tumorales *BRCA* $+/-$, HR est active et va permettre de réparer ces cassures efficacement maintenant ainsi la survie des cellules.

Chapitre II

Place du séquençage haut-débit dans le diagnostic des prédispositions héréditaires au cancer

I- De la découverte de la structure de l'ADN au séquençage du génome humain

C'est en 1953, que James Watson, Francis Crick, Maurice Wilkins et Rosalind Franklin ont démontré que l'ADN a une structure de double hélice et qu'il est composé de quatre bases (Watson & Crick, 1953). Le prix Nobel de physiologie et de médecine en 1962 leur a été attribué pour leurs découvertes qualifiées de plus grande réussite scientifique de leur siècle. Depuis lors jusqu'à nos jours, la biologie a connu différentes révolutions telles que l'invention du séquençage Sanger, de la PCR (Polymerase Chain Reaction) et plus récemment du séquençage haut-débit qui ont permis d'accéder à la séquence de notre génome, à sa variabilité et à l'impact de cette variabilité en génétique humaine.

En 1977, Frederick Sanger a lancé la technologie de séquençage de l'ADN basée sur la méthode des didésoxyribonucléotides ou de terminaison de chaîne d'ADN (**Séquençage Sanger**) qui a révolutionné la biologie et qui est aujourd'hui universellement employée pour séquencer l'ADN (Sanger et al., 1977b). Cette méthode qui a permis le séquençage du premier génome, celui d'un virus bactérien phi X174 (Sanger et al., 1977a) repose sur l'utilisation des nucléotides terminateurs de séquence : les didésoxyribonucléotides (ddNTP). Ces nucléotides ont la particularité de ne pas posséder d'extrémité 3' hydroxyl (OH) libre, empêchant de poursuivre l'élongation. Utilisés en mélange avec des dNTPs classiques, ces dNTPs permettent d'obtenir des fragments marqués de tailles variables qui seront séparés sur un gel de polyacrylamide. La longueur du fragment associée au marquage du nucléotide permet la lecture de la séquence. Cette stratégie a ensuite bénéficié de l'invention de la PCR, du développement de l'électrophorèse capillaire, et de la possibilité de marquage fluorescent permettant de simplifier la partie séparative et analytique. Ainsi cette méthode est capable de générer une longueur de séquence allant jusqu'à 1 kb. Aujourd'hui, les séquenceurs automatisés les plus récents (ex : Séquenceur ABI 3730XL, 3730xl DNA Analyzer, Applied Biosystems, Foster City, CA) utilisant un système de détection in situ pendant l'électrophorèse sont capables de séquencer jusqu'à 2100 kb (kilobases) par jour, par fragments de 400 à 900 paires de bases.

Cette méthode de séquençage de première génération a joué un rôle historique en permettant de réaliser le **projet génome humain**, projet phare de la génétique qui a eu pour but de décrypter l'ensemble du génome humain. Ce plus grand projet biologique collaboratif au monde porté par le Pr Francis Collins a été lancé en 1990 à partir de l'ADN de **100 individus volontaires prélevés**. Après des années de travail collaboratif, et plus de 3 milliards de dollars, une première version du génome humain (3,2 milliards de paires de

bases) fut publiée dans la revue *Nature* en 2001 (Lander *et al.*, 2001) avec la publication d'une cartographie quasi-complète des gènes suivie d'une deuxième version modifiée en 2004 (International Human Genome Sequencing Consortium, 2004). Ce projet qui représentait un effort véritablement international avec une quantité de collaboration sans précédent a rendu possible l'identification de nouveaux gènes associés à différentes maladies.

Cependant, le génome humain de référence généré par ce projet correspond à un mélange de séquences de plusieurs dizaines d'individus. En 2004, le coût estimé pour réaliser un nouveau génome était encore supérieur à 10 millions de dollars. La nécessité de pouvoir séquencer un génome individuel à un coût plus raisonnable a motivé le développement de nouvelles technologies de séquençage capables de séquencer des milliers de fragments d'ADN en parallèle.

II- La révolution du séquençage à haut-débit

En 2004, l'apparition du premier séquenceur à haut-débit commercial (le séquenceur 454) de la société Roche (Roche diagnostics, Almere, The Netherlands) s'est révélée être une vraie révolution dans le domaine du séquençage de l'ADN. Grâce à cette machine, le premier génome complet individuel qui était celui de James Watson a pu être réalisé en 2 mois. Ces automates sont appelés des séquenceurs à haut-débit ou des séquenceurs de nouvelle génération souvent regroupés sous l'acronyme NGS pour *Next Generation Sequencing*, ou HTS (*High-Throughput Sequencing*) permettent de réaliser le séquençage de millions de fragments d'ADN en parallèle. C'est le développement et le rapprochement de trois domaines disciplinaires basés sur la microfluidique, la nanotechnologie et l'informatique, le développement de ces nouveaux types de machines capables de réaliser plusieurs millions qui a permis ce saut technologique (**Figure 5**).

En 2007 sont apparues sur le marché des machines dotées de débits de 50 à 1 000 fois supérieurs aux techniques de première génération. Aujourd'hui, une seule machine peut séquencer 1 Térabases (Tb) en moins d'un jour, et un volume de données équivalent à plusieurs génomes humains complets ou à plus d'une centaine d'exomes en moins de 48 heures (ex : NovaSeq, Illumina, Eindhoven, The Netherlands) (**Figure 6**). Ce développement technologique s'est également accompagné d'une diminution drastique des coûts de séquençage (**Figure 7**). Grâce à une réduction spectaculaire des prix, le NGS a rapidement mis le séquençage du génome à la portée de laboratoires, et l'objectif initial de séquencer un génome humain pour moins de 1000 US \$ a été atteint il y a quelques années. Une illustration de ces avancées est le projet 1 000 Génomes, débuté en 2008, qui avait pour objectif de séquencer le génome de 2500 personnes, afin d'identifier les spécificités de chacun et rendre possible une médecine personnalisée, avec des traitements adaptés aux spécificités génétiques de chacun (1000 Genomes Project Consortium et al., 2010).

Depuis, de nombreux projets nationaux ou internationaux ont été développés dans le but de séquencer un maximum de génomes en population générale. Parmi ceux-ci, on peut citer le projet GnomAD (Genome Aggregation Database) qui regroupe plus de 15 496 génomes et 123 136 exomes (<http://gnomad.broadinstitute.org>).

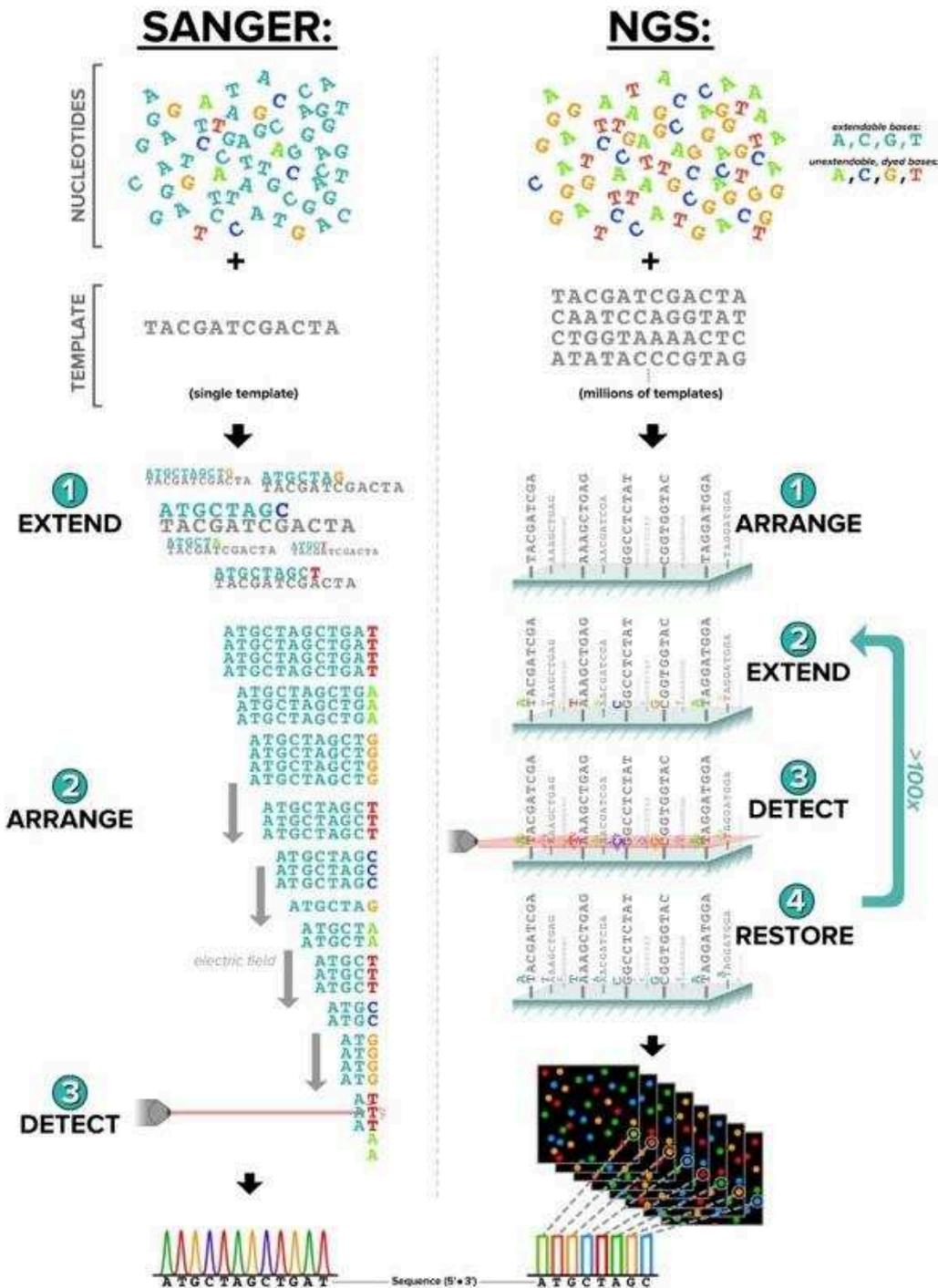


Figure 5 : Comparaison entre le séquençage traditionnel de Sanger et la technologie de séquençage de nouvelle génération (D’après Price *et al.*, 2018)

Le séquençage Sanger permet de séquencer un fragment alors que le séquençage à haut-débit fait référence à un ensemble de technologies permettant le séquençage en parallèle de millions de fragments d’ADN.

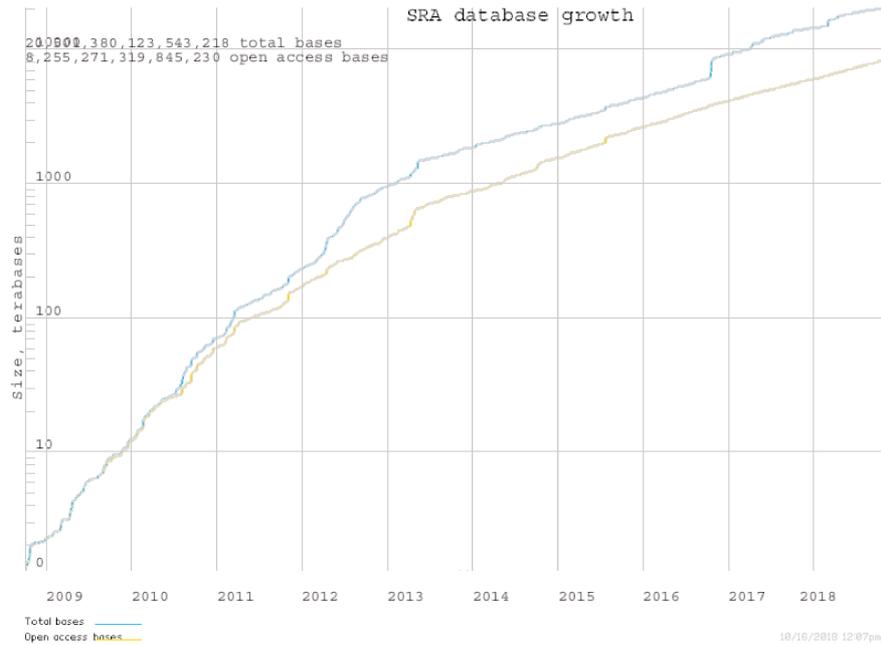


Figure 6 : Graphique illustrant la capacité de séquençage mondiale en térabases en fonction du temps (Données issues du SRA (*Sequence Read Archive*) à la date du 2 Août 2018, <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>)

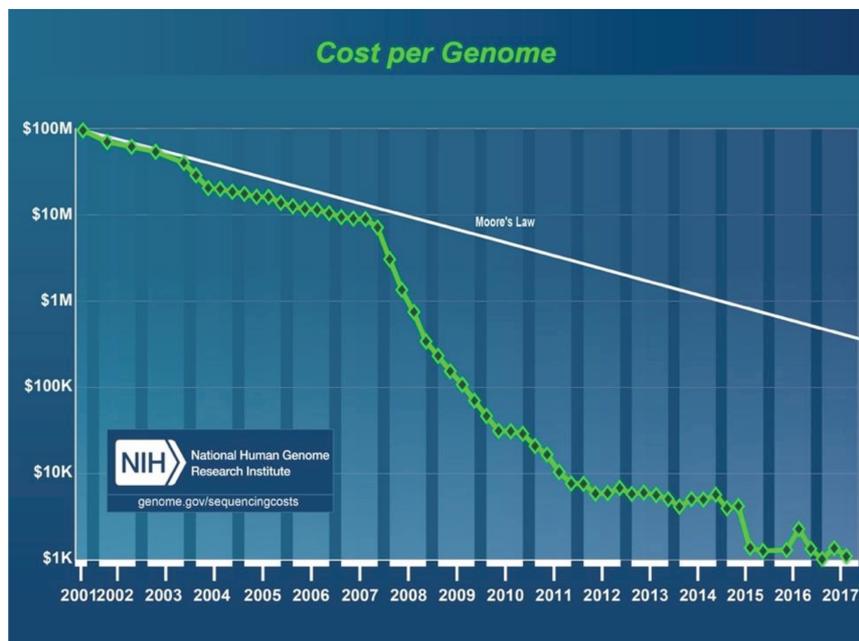


Figure 7 : Graphe illustrant l'évolution du coût de séquençage d'un génome complet en fonction du temps (<https://www.genome.gov/sequencingcostsdata/>, à la date du 25 Avril 2018) (Données issues du *National Human Genome Research, Institute Genome Sequencing Program GSP*)

III- Principe du séquençage NGS

Le séquençage à haut-débit consiste en un séquençage simultané de milliard de fragments différents. Plusieurs étapes sont nécessaires à ce séquençage massif. La première étape consiste à établir une librairie d'ADN compatible avec le séquençage. Cette librairie est ensuite amplifiée de façon clonale puis séquencée. La phase finale consiste en l'analyse bio-informatique des résultats avant que le résultat devienne utilisable par les biologistes.

A partir des molécules d'acides nucléiques ou ribonucléiques, il est nécessaire d'obtenir une librairie d'ADN double brin de petite taille contenant les séquences adaptatrices essentielles au séquençage.

1- Préparation des librairies

La première étape consiste à produire des millions de petits fragments d'ADN génomique ou d'ADNc à partir du matériel du départ par une fragmentation aléatoire. Cette fragmentation peut être obtenue par sonication (traitement aux ultrasons), par l'utilisation de transposases ou par l'utilisation de nucléases non spécifiques pour donner des fragments dont la taille est compatible avec le séquenceur (petits fragments de 200 à 300 nucléotides en moyenne pour les séquenceurs de 2^{ème} génération).

Il est nécessaire pour pouvoir amplifier tous ces fragments différents en une seule réaction de leur lier des séquences adaptatrices aux extrémités. Ces séquences adaptatrices contiennent les séquences complémentaires des amorces de PCR et de séquençage. Elles peuvent également contenir plusieurs éléments :

- un index ou code barre correspondant à un échantillon et permettant de séquencer plusieurs échantillons sur le même run de séquence.

- une autre succession dégénérée de 8 à 12 paires de bases correspondant à des identifiants de molécules uniques (UMI, *Unique Molecular Identifier*) et permettant de s'affranchir des erreurs de PCR et de séquençage.

Avec les technologies de séquençage, ont évolué les technologies de préparation des librairies. Il est maintenant possible de choisir les régions que l'on veut séquencer pour diminuer les coûts de séquençage ainsi que le temps d'analyse et les besoins de stockage informatique.

Deux grands types d'enrichissement sont actuellement disponibles pour répondre à ce besoin. L'enrichissement par PCR multiplexe lorsque les régions à enrichir sont de l'ordre d'un à une dizaine de gènes et l'enrichissement par capture en phase liquide lorsque la taille de la région est plus importante (séquençage exomique ou de grands panels de gènes).

- **Enrichissement par capture** consiste à capturer les fragments d'intérêt par l'utilisation des sondes complémentaires de ces régions. Après hybridation des sondes sur les régions d'intérêt, les duplex sont récupérés sur des billes magnétiques qui peuvent être recouvertes de streptavidine possédant une très forte affinité pour la biotine qui marque les sondes (*The Agilent Technologies SureSelect™, California, United States*). Après différents lavages, ces fragments d'intérêt sont élués et amplifiés pour pouvoir être dosés de façon précise (Bashiardes *et al.*, 2005) (**Figure 8**).
- **Enrichissement par PCR multiplexe** : pour construire des libraires avec la technologie de PCR multiplexe les étapes de fragmentation et d'ajout d'adaptateurs décrits ci-dessus ne sont pas nécessaires. En effet, les séquences adaptatrices seront apportées par les différents primers de PCR utilisés dans cette réaction de PCR multiplexe destinée à sélectionner et amplifier les régions cibles. Pour des considérations de coût et de facilité, les séquences adaptatrices sont généralement ajoutées en 2 fois lors de la PCR multiplexe spécifique des séquences cibles puis dans la deuxième PCR universelle (**Figure 9**).

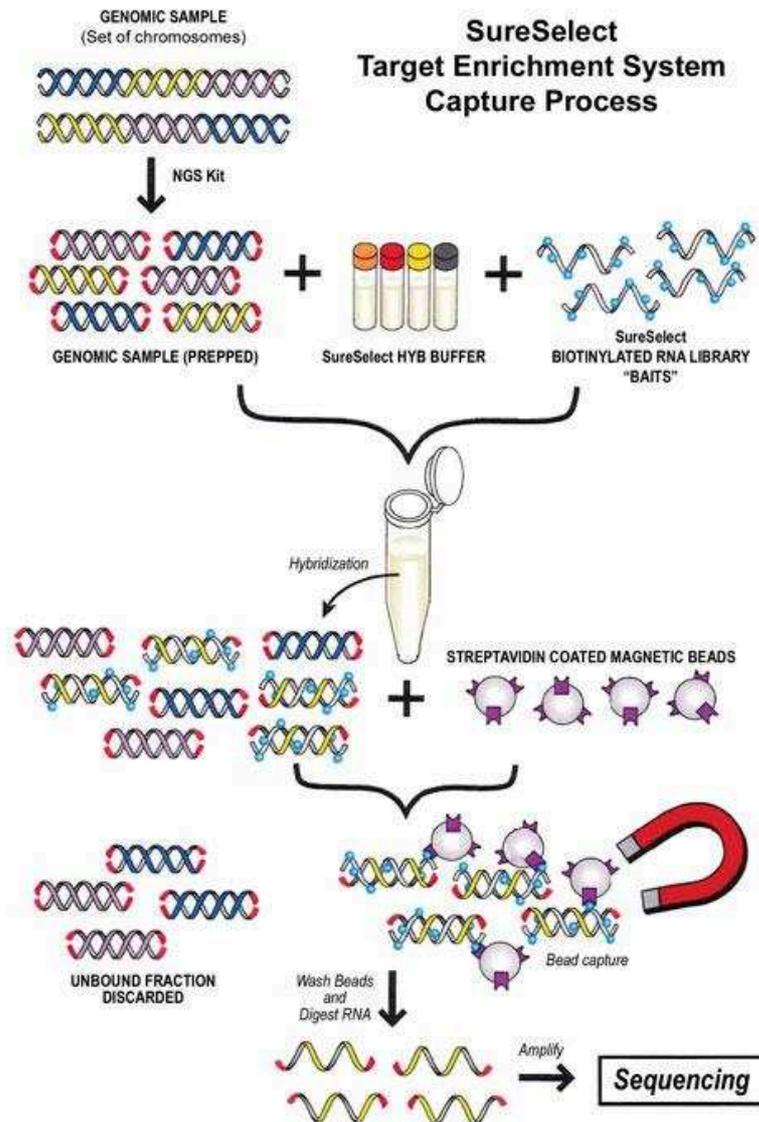


Figure 8 : Enrichissement par capture en phase liquide des régions d'intérêt – Technologie SureSelect (Agilent)

Lors d'un enrichissement par capture, l'ADN génomique est coupé en fragments d'environ 200 bases. Ensuite, un système de sondes ARN biotinylées complémentaires aux régions d'intérêts sont ajoutées dans le milieu pour capturer les régions cibles. Après une incubation, les hybrides sondes ARN/fragments ADN d'intérêt sont capturés à l'aide de billes magnétiques recouvertes de streptavidine, fortement affines pour la biotine. Grâce à un aimant, seuls ces complexes sont retenus lors des phases de lavages. Les régions non capturées sont éliminées pour ne conserver que les fragments d'intérêt en vue du séquençage.

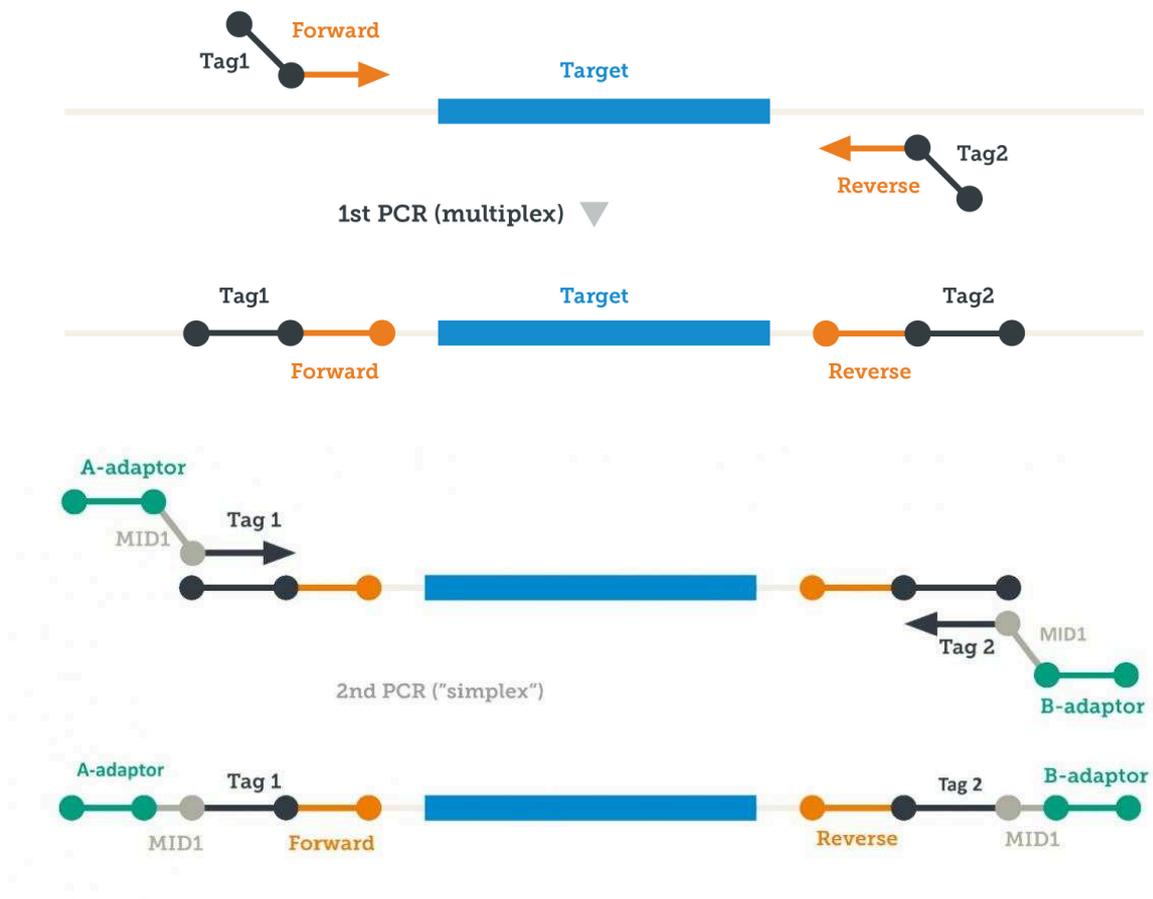


Figure 9 : Représentation schématique du principe d'enrichissement par PCR multiplexe (Source : <https://www.agilent.com>)

Lors d'un enrichissement par PCR multiplexe, les régions d'intérêt sont amplifiées de façon simultanées en utilisant des amorces spécifiques allongées de séquences adaptatrices (Tag 1 et 2). Ensuite des adaptateurs spécifiques (A-adaptor et B-adaptor) sont ajoutés aux produits de la première PCR obtenus ainsi qu'un code barre permettant d'identifier le patient (MID1) à l'aide d'une PCR avec un seul couple de primer.

2- Amplification clonale

Cette étape consiste à amplifier tous les fragments de la librairie établie de façon simultanée tout en gardant la clonalité ; un fragment doit générer un clone de molécules identiques qui doivent être restés séparés des autres clones. La compagnie **Life Technologie** (Californie, États-Unis) et la compagnie **Illumina** sont actuellement les deux leaders du marché des séquenceurs de 2^{ème} génération. La méthode d'**amplification** nécessaire pour l'étape de séquençage peut varier en fonction de la plate-forme choisie.

➤ Pour l'amplification clonale, la société **Life Technologie** a opté pour la **PCR en émulsion** (emPCR) sur ses plateformes PGM Ion Torrent et Proton. La PCR se fait sur un support solide (une bille microscopique, ISP Ion Sphere Particule). La molécule d'ADN est immobilisée sur la bille via son adaptateur. Les billes sont capturées dans des gouttelettes aqueuses contenant l'ensemble des réactifs nécessaires à la réaction où se réalise la PCR afin d'obtenir des amplicons à la surface de chaque bille. La dilution de la librairie et des billes est faite pour avoir au maximum une bille et une molécule par gouttelette assurant ainsi la clonalité. Ces ISP amplifiées sont ensuite chargées sur une puce ou chip pour que les fragments d'ADN clonés soient directement séquencés (**Figure 10, A**).

➤ La plate-forme Illumina utilise l'amplification sur un support solide qui est généralement une lame de verre (*flowcell*) : Chaque fragment de la librairie se lie à la surface de la puce via son adaptateur. Les fragments sont amplifiés par PCR clonale par formation de ponts grâce aux séquences adaptatrices, afin d'obtenir des groupes d'amplicons ou clusters à la surface de la puce. Puis, ces fragments clonés sont directement séquencés sur la *flowcell*. Là encore, la clonalité dépend de la séparation physique des clusters sur la *flowcell* et donc de la densité de chargement (**Figure 10, B**).

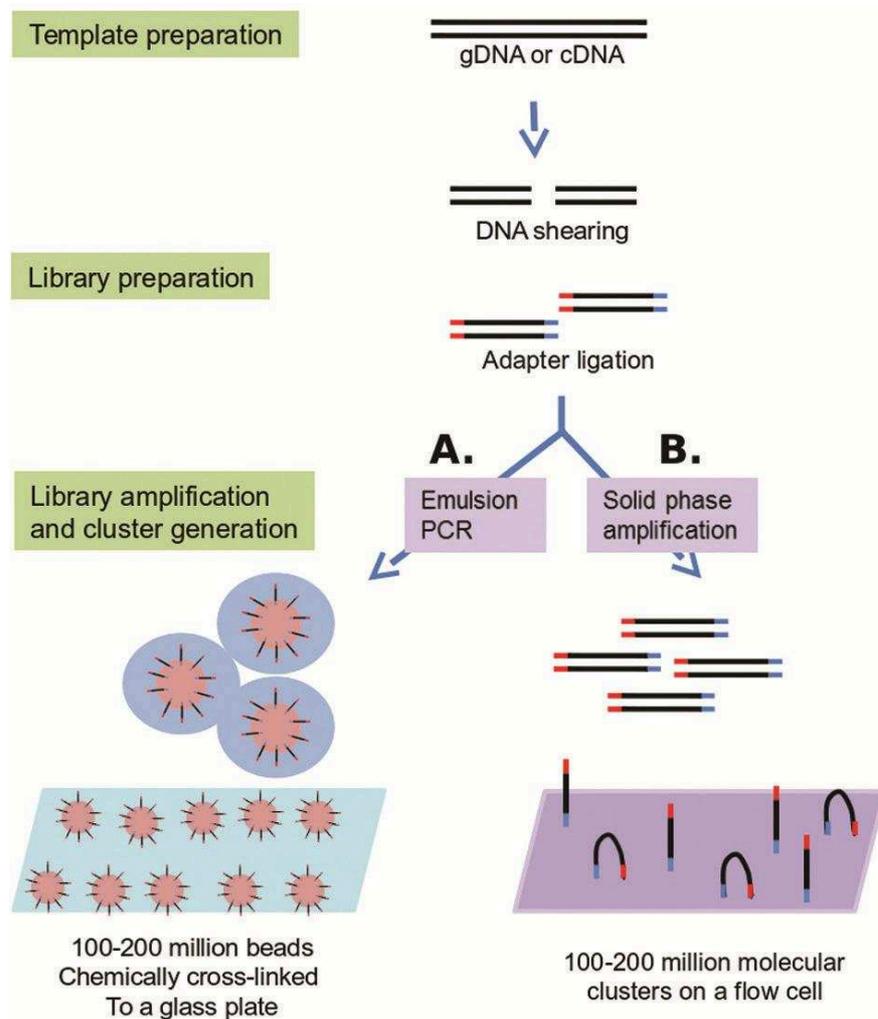


Figure 10 : Schéma montrant les deux méthodes d'amplification clonale des librairies (D'après Bahassi & Stambrook, 2014)

(A) Amplification à l'aide d'une PCR en émulsion ; (B) Amplification en pont en phase solide

A : Une émulsion huile-aqueuse est créée pour encapsuler les complexes billes-ADN en gouttelettes aqueuses uniques. L'amplification en PCR est réalisée au sein de ces gouttelettes contenant chacune plusieurs milliers de copies de la même séquence du fragment d'ADN original. Les billes sont ensuite liées chimiquement à une lame de verre.

B : Les fragments d'ADN flanqués de leurs adaptateurs universels sont distribués sur la flowcell. Des séquences complémentaires attachées sur la flowcell vont s'hybrider aux adaptateurs. Après la distribution des éléments nécessaires à l'amplification, une production de millions de clusters séparés spatialement après une amplification par pont va être ainsi réalisée.

3- Séquençage à haut-débit

A- Séquençage de 2^{ème} génération

Après l'étape d'amplification clonale des fragments d'ADN, les échantillons sont à proprement séquencés. Plusieurs technologies de séquençage à haut-débit sont aujourd'hui accessibles aux laboratoires. Les séquenceurs de deuxième génération permettent le séquençage en parallèle de courts fragments d'ADN avec des lectures allant de 50 à 300 pb.

Il existe aujourd'hui deux approches distinctes de séquençage se focalisant sur la lecture de fragments ADN courts :

a- **Le séquençage par synthèse, Sequencing By Synthesis, SBS (Illumina, Ion Torrent) :**

Dans les approches par synthèse, une polymérase va former le brin complémentaire de la séquence d'ADN à lire, l'incorporation d'un nucléotide étant signalée par une émission lumineuse (fluorophore) ou par un changement dans la concentration ionique du milieu réactionnel. Il est possible de catégoriser le séquençage par synthèse en deux sous-ensembles :

- La technologie CRT (Cyclic Reversible Termination) : basée sur l'incorporation étape par étape de nucléotides fluorescents terminateurs de séquence réversibles pour le séquençage de l'ADN. Chacune des 4 bases émet un signal de fluorescence unique lorsqu'elle est ajoutée au brin en cours de synthèse. La mesure de la fluorescence est ensuite convertie en données de séquence. Les reads obtenus sont très courts, mais la technique est robuste et relativement bon marché. Cette méthode développée par Solexa et utilisée par la technologie Illumina® peut être utilisée pour séquencer l'ADN de n'importe quel organisme, fournissant des informations précieuses en réponse à presque n'importe quelle question biologique (**Figure 11, a**).

- La technologie SNA (Single Nucleotide Addition) basée sur la quantification d'un sous-produit de la réaction d'élongation. Il peut s'agir du pyrophosphate inorganique (PPi) avec la technique de pyroséquençage ou des protons H⁺ libérés par la polymérase lors de l'ajout d'une base avec la technique de Proton.

- Le pyroséquençage : Dans cette réaction de séquençage la libération du pyrophosphate résultant de l'incorporation des nucléotides est utilisée dans une

série de réactions chimiques entraînant la génération de flashes lumineux. L'émission de lumière est détectée par une caméra qui enregistre la séquence appropriée du cluster (Nyrén, 2007). Ce cycle se poursuit jusqu'à ce que la réaction de séquençage soit terminée. Cette technique assez robuste utilisée par la plateforme **454** (Life Science –Roche) présente malheureusement un taux d'erreur élevé dans les séquences d'homopolymères (Margulies *et al.*, 2005) (**Figure 11, b**).

- Séquençage par semi-conducteur d'ions : La technologie **Ion-Torrent** (Thermo Fischer Scientific) utilise des semi-conducteurs et offre un autre type de plateforme. Les quatre nucléotides sont successivement apportés. Lors de leur incorporation par l'ADN polymérase, le nucléotide va modifier le pH du milieu réactionnel en libérant un proton (H^+). Chaque bille est située directement au-dessus d'un transistor semi-conducteur qui est capable de détecter les changements du pH de la solution. Ensuite la différence de potentiel due au H^+ libérés est traduite en données de séquence. Cette approche est très bon marché et rapide, mais moins fiable en raison de difficultés techniques et des taux d'erreur élevés sur les segments homopolymériques de nucléotides (supérieur à 6-8 pb) (Goodwin *et al.*, 2016) (**Figure 11, c**).

b- Le séquençage par ligation, Sequencing By Ligation, SBL :

Le séquençage par ligation est différent des deux autres méthodes car il n'utilise pas d'ADN polymérase pour incorporer les nucléotides. Au lieu de cela, il repose sur de courtes sondes oligonucléotidiques qui sont liguées les unes aux autres. Dans les approches par ligation, une séquence ADN sonde liée à un fluorochrome va pouvoir s'hybrider aux fragments d'ADN à séquencer. Cette sonde hybridée pourra être liguée au primer de séquence puis aux autres sondes. La couleur de l'émission lumineuse du fluorochrome va indiquer l'identité de la ou des bases liguées. Cette méthode utilisée par la plateforme **SOLiD** utilise un procédé de séquençage par ligation complexe qui est très robuste, mais lent et cher (**Figure 11, d**).

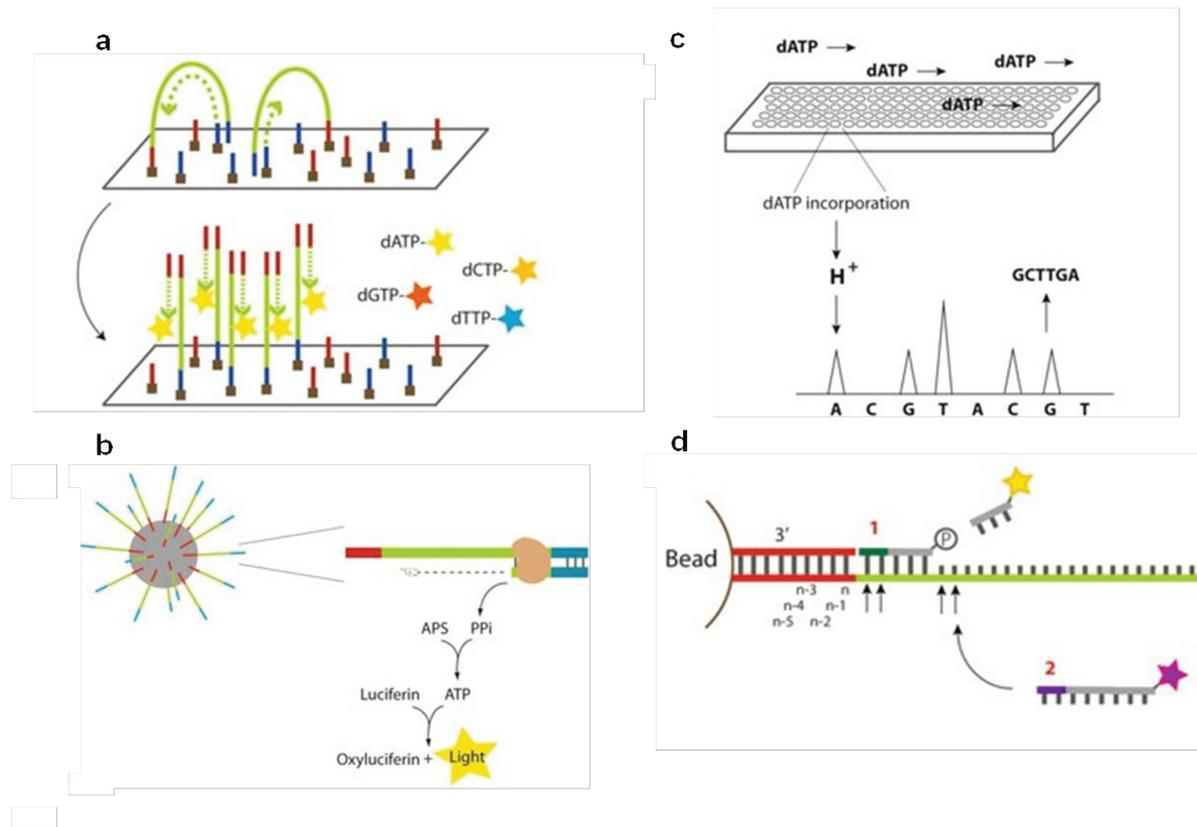


Figure 11 : Aperçu des technologies de séquençage de deuxième génération (D'après Wong, 2013)

(a) séquençage par synthèse (Illumina) ; (b) pyroséquençage (Roche 454) ; (c) séquençage par synthèse et puce à semi-conducteur (Ion Torrent, Life Technologies) ; (d) séquençage par ligation (SOLiD, Life Technologies).

Chaque plateforme a donc ses propres caractéristiques avec des avantages et des inconvénients (Boland et al., 2013). Pour différentes raisons et notamment pour des raisons de précision de séquençage, la société Illumina est aujourd'hui en position leader sur le marché des séquenceurs à haut-débit de 2^{ème} génération.

B- Séquençage de troisième génération

L'impossibilité de séquencer de longues molécules d'ADN complexes se trouvant dans notre génome est l'une des limites du séquençage à haut-débit deuxième génération. Pour pallier à ce défaut, plusieurs sociétés ont développé des approches artificielles qualifiées de longues lectures synthétiques (synthetic long read) permettant de mettre en évidence des variants structuraux de grande taille de l'ADN telles que les variations du nombre de copies, les inversions et les translocations et de mieux séquencer les régions fortement répétées. Ce type de séquençage facilite ainsi l'assemblage *de novo* de nouveaux génomes, et permet également de reconstruire des haplotypes complets ou de lire des transcrits ARNm entiers. Cette approche des longs reads synthétiques est basée sur l'utilisation de séquenceurs de 2^{ème} génération générant des lectures courtes pour construire des reads longs *in silico*. Le protocole implique une fragmentation mécanique initiale de l'ADNg (jusqu'à 10 kb) en fragments de grande taille. Ces fragments vont être dilués sur des plaques à plus de 350 puits de sorte que chaque puits contienne au maximum une molécule soit environ 1 à 2 % du génome. Ensuite l'ADN est de nouveau fragmenté dans chaque puits et des codes-barres différents sont ajoutés à l'intérieur de chaque puits. Après séquençage, les données sont ensuite rassemblées par code-barres (démultiplexées) tout en sachant que les fragments présentant le même code-barre sont dérivés du même grand fragment original. Les lectures longues synthétiques sont ensuite assemblées à partir des lectures courtes. Actuellement, deux systèmes sont disponibles pour générer des longs reads synthétiques : le système Moleculo d'Illumina et le système 10X Genomics.

Cette approche reste cependant dépendante des séquenceurs de 2^{ème} génération et donc d'une étape d'amplification clonale qui peut être à l'origine d'artéfacts de PCR et de surreprésentation de région du génome plus facilement amplifiable que d'autres. Pour dépasser ce besoin, une troisième génération de séquenceurs a été développée. Ces séquenceurs troisième génération sont capables de générer des reads longs (jusqu'à 20 kb ou même 30 kb) et ne nécessitent plus d'amplification clonale. Les deux leaders de cette approche à l'heure actuelle sont Pacific Biosciences Technologies et Oxford Nanopore Technologies.

La technologie développée par Pacific Biosciences se base sur une lecture en temps réel d'une molécule unique (SMRT, Single Molecule Real Time). Le PacBIO utilise des nanostructures appelées zero-mode waveguides ou ZMWs qui contiennent un volume de l'ordre du zeptolitre (10³ nm³ soit 10⁻²¹ L) avec une polymérase fixée de façon covalente au

fond de la nanostructure. Lors du séquençage d'une molécule unique, les nucléotides fluorescents qui entrent dans la cavité de l'ADN polymérase pour être incorporés sont détectés par le plus petit microscope jamais développé au fur et à mesure de leur incorporation. Ce signal fluorescent est ensuite retranscrit sous forme de bases. Cette technologie permet de séquencer des reads supérieurs à 20 kb avec des temps de séquençage très courts (Eid et al., 2009).

La seconde approche développée par Oxford Nanopore Technologies (ONT) est le séquençage par nanopores. Dans cette technologie, un nanopore protéique est fixé dans une membrane synthétique polarisée résistante à l'électricité. Un courant ionique est fait passer à travers le nanopore en établissant une tension à travers cette membrane. La molécule d'ADN à séquencer est faite passer au travers d'un nanopore. Le courant est modifié lorsque les bases G, A, T et C traversent le pore dans différentes combinaisons. Le suivi dans le temps des modifications de la membrane permet de déterminer la séquence nucléotidique (Clarke et al., 2009; Loose et al., 2016).

Les nouvelles technologies de séquençage en temps réel présentent cependant l'inconvénient de taux d'erreurs de séquençage élevés par rapport aux séquençages de 2^{ème} génération. Cependant, cet inconvénient semble se résoudre par la possibilité de séquencer une même molécule plusieurs fois (en boucle) pour lisser les erreurs de séquençage et obtenir une séquence consensus plus fidèle.

4- Analyse bio-informatique

Le séquençage à haut-débit génère donc des millions de courtes lectures appelés « *reads* » qui sont déduites des données brutes par une étape bioinformatique appelée « *Base calling* ». Ces lectures sont rassemblées dans un fichier FastQ. En cas de multiplexage, chaque *read* est réassigné au bon échantillon grâce au code barre utilisé lors de l'étape de **démultiplexage**. Les lectures obtenues pour un échantillon sont ensuite alignées sur un génome de référence lors du processus d'**alignement**. Les fichiers d'alignement sont nommés BAM (*Binary Alignment Map*) et peuvent être visualisés à l'aide de *Genome Browsers* tels que IGV (*Integrative Genomic Viewer*, <http://software.broadinstitute.org/software/igv/>) ou UCSC (UCSC, *University of California Santa Cruz*, <https://genome.ucsc.edu>). Dans le cas du

séquençage d'un génome pour lequel il n'existe pas de séquence référence, il est possible d'effectuer un **alignement de novo**.

Une fois les données de séquence alignées, elles pourront être utilisées afin de mettre en évidence les variants de la séquence nucléotidique par rapport à la séquence de référence à travers l'étape bioinformatique **d'identification des variants** (*Variant Calling*) qui génère des fichiers vcf (*variant call format*). Cette étape peut être réalisée avec différents outils comme l'haplotypcaller de GATK ou l'outil VarScan en fonction du type de mutation recherché (mutations à l'état hétérozygote ou homozygote ou mutations présentes à bas bruit) (McKenna *et al.*, 2010 ; Koboldt *et al.*, 2012). Le meilleur compromis entre spécificité et sensibilité est généralement obtenu en combinant plusieurs outils différents. La liste des variants détectés est ensuite annotée pour rassembler toutes les informations disponibles sur ce variant lors du **processus d'annotation** (nomenclature en c., impact fonctionnel du variant, fréquence en population générale, prédictions bioinformatiques de l'impact du variant etc.).

Cette analyse s'accompagne également d'une **analyse qualité** qui permet de suivre le bon déroulement du séquençage et de la préparation des librairies et qui permet notamment d'évaluer la profondeur moyenne de lecture ainsi que la qualité moyenne des bases lues. Enfin, dans le cadre d'une activité diagnostique, il est essentiel de disposer des **bases de données** adéquates pour pouvoir stocker ces informations et pouvoir réaliser des requêtes sur les variants, mais également pour pouvoir échanger des données d'un laboratoire à l'autre.

Sur les données générées par DNA-Seq, il est possible d'identifier les variants ponctuels mais également les CNV (*Copy Number Variations*) en se basant notamment sur l'analyse de la profondeur de lecture, sur la présence de lectures chimériques contenant les points de cassure des remaniements ou en se basant sur l'analyse de la taille théorique des inserts lors de séquençage en paired-end (Alkan *et al.*, 2011).

IV- Diagnostic des formes héréditaires de cancer et NGS

Aujourd'hui, grâce aux évolutions des technologies et à la baisse des coûts de séquençage, le NGS est une technologie largement utilisée avec des applications allant de la biologie fondamentale à la médecine en passant par l'agriculture et l'agronomie. La miniaturisation des séquenceurs NGS avec le développement de « séquenceurs NGS de paille » , la diminution drastique des coûts de séquençage et le développement de techniques d'enrichissement permettant de sélectionner les zones du génome à séquencer ont permis une entrée massive des séquenceurs à haut-débit dans les laboratoires de diagnostic moléculaire et un changement radical des pratiques.

Classiquement, les méthodes de diagnostic traditionnelles se concentrent sur l'identification des mutations par approche gène par gène en séquençant les exons de ces gènes par séquençage Sanger, ou en recherchant les remaniements génomiques de grande taille par les techniques de PCR multiplexe semi-quantitative (QMPSF, Quantitative Multiplex PCR of Short Fluorescent fragments) ou d'amplification multiplexe de sondes dépendante d'une ligation (MLPA, Multiplex Ligation-dependent Probe Amplification) permettant de détecter les délétions et les duplications au niveau de ces exons. La démocratisation des séquenceurs à haut-débit a révolutionné la façon dont le diagnostic moléculaire est effectué dans de nombreux laboratoires. En effet, les approches classiques sont en train d'être remplacées par des approches à haut-débit permettant des analyses multi-gènes et permettant de détecter simultanément les SNV (Single Nucleotide Variant) mais également les CNV. Avec les nouvelles technologies, les laboratoires peuvent analyser plusieurs gènes différents en même temps et de façon très rapide et économique par rapport au séquençage Sanger. L'augmentation du débit des analyses, la diminution des coûts et le développement des méthodes d'enrichissement ont permis d'inclure dans les tests diagnostiques, d'autres gènes de prédisposition dont les mutations ont une très faible prévalence dans la population, mais dont les conséquences sont proches de celles des mutations dans les gènes de prédisposition majeurs. Au titre de la recherche, il est également possible d'inclure dans les gènes séquencés des gènes dont l'implication clinique n'est pas encore formellement démontrée. Ces approches diagnostiques à haut-débit peuvent concerner une liste de gènes limitée (de 2 à 500 gènes par exemple) ou bien concerner l'ensemble des 20 000 gènes ou encore l'ensemble du génome humain.

1- Approche diagnostique en panels de gènes cibles

Avec les progrès récents de la technologie NGS, des tests de panels de gènes (ou multi-gènes) ont été introduits depuis 2010 dans le but de faciliter le diagnostic des formes mendéliennes de cancer (Walsh *et al.*, 2010; Pritchard *et al.*, 2012). Ces panels contiennent un ensemble sélectionné de gènes ou de régions de gènes associés à un ou plusieurs phénotypes de cancer et qui sont examinés en même temps dans une même réaction. Ces tests génétiques ciblent généralement un nombre limité de gènes (de 2 à 200) et sont souvent employés dans des stratégies pour lesquelles les résultats d'un test préliminaire auraient conditionné le suivant (séquençage du gène A puis si retour négatif, séquençage du gène B). Ces panels sont soit définis par des sociétés commerciales tels que GeneDX Myriad ou AmbryGenetics, soit personnalisés pour s'adapter aux pratiques d'un laboratoire en incluant des gènes spécifiques d'une forme héréditaire de cancer ou de plusieurs formes héréditaires touchant le même organe. Ces panels peuvent également être adaptés à l'activité diagnostique du laboratoire en regroupant pour des questions de coûts ou d'organisation des gènes impliqués dans plusieurs pathologies différentes.

Le séquençage d'un panel de gènes permet donc un **séquençage rapide et simultané** des gènes impliqués dans une même pathologie. Il permet ainsi de **réduire les délais de rendu au patient** en comparaison avec les approches de séquençage gène par gène. Il permet également **d'inclure un grand nombre de gènes expérimentaux** ayant une utilité clinique potentielle pour un coût supplémentaire peu élevé. Ainsi, dès 2010, différents panels de gènes dédiés au diagnostic des formes héréditaires de cancer du sein et de l'ovaire (HBOC, *Hereditary Breast and Ovarian Cancer*) ont été développés. L'équipe de Mary-Claire King a développé une approche basée sur l'enrichissement par capture de 21 gènes conférant un risque plus élevé de développer un cancer du sein et de l'ovaire (Walsh *et al.*, 2010). Depuis, ce panel a été repris et a évolué pour inclure un plus grand nombre de gènes. Ainsi, en utilisant en routine un panel de gènes contenant des gènes analysés au titre de la recherche, il est possible d'accumuler les données et de trancher quant à l'utilité clinique de séquencer ces gènes candidats. L'équipe Caennaise de notre unité de recherche a ainsi démontré sur plus de 5000 patients l'utilité diagnostique des gènes *TP53*, *PALB2* (Partner and Localizer of BRCA2), *RAD51C* et *D* dans le cadre de la prédisposition aux cancers du sein et de l'ovaire (Castéra *et al.*, 2014, 2018).

L'inclusion en un seul et même panel de gènes impliqués dans plusieurs formes héréditaires de cancer permet d'adresser la question d'un éventuel **chevauchement**

phénotypique de ces syndromes. Notre équipe a ainsi développé pour le diagnostic du cancer colorectal héréditaire une stratégie de séquençage basée sur la capture rapide des séquences exoniques et introniques de dix gènes impliqués dans les formes mendéliennes de cancer colorectal. Cette stratégie permet, en une seule étape, de détecter tous les types d'altérations du gène, y compris les variants de structure, et fournit un taux élevé de détection de variants délétères, optimisant ainsi le diagnostic des formes héréditaires de cancer (Baert-Desurmont et al., 2018).

Le séquençage à haut-débit de panels de gènes à façon offre également la possibilité de séquencer plus que les exons et **d'inclure dans le panel des régions régulatrices** telles que les promoteurs et les régions 5' et 3' UTR (Untranslated Regions) ainsi que les introns. L'inclusion des régions introniques permet également de faciliter la détection des réarrangements de grande taille sur la base des données générées par le NGS (Baert-Desurmont et al., 2018). En utilisant une profondeur de séquençage adaptée, cette approche en panel de gènes permettra à un coût raisonnable de détecter les **mutations en mosaïques** des gènes de prédisposition au cancer qui peuvent être difficiles à détecter par les techniques classiques (Renau-Petel et al., 2018).

Ces panels de gènes peuvent être utilisés plus largement pour étudier les bases moléculaires à l'origine de cas précoces de cancer pour lesquels aucune mutation délétère n'a pu être mise en évidence dans les gènes connus. Récemment, notre équipe a développé en collaboration avec le site de Caen, un panel de 201 gènes impliqués dans cancer. Ce panel a été appliqué à des cas sélectionnés de cancers évocateurs d'un déterminisme génétique du fait de leur précocité sans mutation détectable des gènes connus de prédisposition génétique au phénotype considéré (données non publiées).

Cette stratégie offre donc une alternative très intéressante par rapport aux techniques classiques. Elle peut être appliquée sur différents types d'échantillons (paraffinés, ADN circulant) avec des quantités d'ADN de départ faibles (50 ng en moyenne pouvant descendre jusqu'à 1 ng). La quantité de données générées est plus facile à gérer, stocker et interpréter par rapport à des approches plus larges telles que le séquençage du génome entier. En revanche, ces panels restent limités en termes de nombre de gènes inclus et la détection des mutations causales reste limitée aux gènes inclus, correspondant à l'état des connaissances sur la pathologie à un moment donné.

2- Approche de séquençage de l'exome entier

Aujourd'hui avec l'amélioration et la baisse des coûts des kits d'exomes, l'approche de séquençage d'exome devient en théorie compatible avec la routine diagnostique. Cette approche consiste à séquencer les exons entiers ou les régions codantes de presque tous les gènes connus qui constituent environ 23 000 gènes, soit 2 % du génome (62 Mb d'information génétique), mais qui contiennent environ 58 % des mutations mises en cause dans les maladies génétiques (Choi et al., 2009). En effet, 85 % des maladies Mendéliennes s'expliqueraient par des altérations présentes dans les régions codantes du génome. Pour cette raison, séquencer l'exome est un outil intéressant pour identifier les bases moléculaires de maladies Mendéliennes telles que les cancers héréditaires.

Cette méthode offre l'avantage de séquencer tous les gènes et de ne plus être restreint à un ensemble de gènes comme cela est le cas dans les panels de gènes. Cette propriété est particulièrement intéressante dans le domaine du diagnostic des prédispositions héréditaires aux cancers, car elle permet d'explorer des gènes qui auraient été criblés en seconde intention en cas de résultat négatif et des gènes qui n'auraient pas été séquencés du tout, au sein d'une même analyse. Ce test peut donc révéler des variants dans des gènes qui n'ont pas encore été impliqués en génétique des cancers et qui deviennent des gènes candidats. Les études de Whole-Exome Sequencing (WES) ont déjà permis d'identifier un certain nombre de gènes causant des syndromes de cancer rares et ont contribué à la compréhension de la génétique des types de cancer plus communs. Par exemple, l'identification des mutations du gène SMARCA4 (Switch/sucrose non-fermentable related, Matrix Associated Actin dependent Regulator of Chromatin subfamily A, member 4) dans les carcinomes hypercalcémiques de l'ovaire a permis une percée dans la compréhension des bases génétiques de cette maladie (Lavrut et al., 2016). De plus cette stratégie a permis de révéler des rôles jusqu'alors inconnus pour certains gènes liés au cancer. Par exemple, le gène PALB2 qui a été initialement impliqué dans la prédisposition au cancer du sein a été impliqué comme gène de prédisposition au cancer du pancréas par des analyses exomiques (Jones et al., 2009).

Il existe cependant quelques limites au séquençage exomique. Les régions d'intérêt étant plus grandes, les données de couverture et de profondeur sont en général moins bonnes que celles des approches en panels de gènes compte-tenu du coût important de séquençage. La détection des mutations en mosaïque est donc plus compliquée. De plus, le nombre des variants détectés est sensiblement plus élevé avec plus de 20 000 variants en moyenne par individu, complexifiant ainsi l'analyse et augmentant le nombre de variants de signification biologique inconnue à interpréter. Le volume de données générées par exome est plus important,

ce qui nécessite des infrastructures adaptées pour le calcul et le stockage informatique de ces données ce qui représente aujourd'hui un énorme challenge pour les services informatiques des différents hôpitaux en charge de ces analyses moléculaires. Il est à noter qu'une analyse d'exome peut également révéler des mutations délétères dans des gènes impliqués dans des pathologies autres que la prédisposition au cancer. Ces découvertes incidentelles pour lesquelles il est possible de proposer une action bénéficiant au patient doivent être rapportées et il est donc essentiel dans ce contexte de repenser les consentements éclairés qui seront signés par les patients.

Si l'approche Exome reste peu adaptée à la routine diagnostique à l'échelle d'un laboratoire, elle devient applicable à l'échelle d'un réseau. En effet, l'organisation proposée par le Plan France Médecine Génomique propose que le séquençage des exomes soit réalisé sur des plateformes nationales et que les analyses soient réalisées sur des centres de calcul concentrant les ressources informatiques nécessaires. Avec cette organisation, l'exome diagnostique deviendrait donc une réalité y compris pour les formes Mendéliennes de cancer.

3- Approches de panel *in silico* sur séquençage exomique

Le séquençage utilisant un panel *in silico* offre une alternative entre le séquençage de panels de gènes et le séquençage d'exome. Après le séquençage exomique ou tout génome, les résultats sont filtrés bioinformatiquement en utilisant un panel *in silico* correspondant à une liste de gènes d'intérêt. Cette approche permet aux laboratoires de bénéficier du WES ou du WGS (Whole Genome Sequencing), tout en s'affranchissant des données difficilement interprétables (variants intergéniques ou dans des gènes sans lien avec la pathologie étudiée, incidentalomes). En fonction de l'évolution des connaissances sur la pathologie, ces données de séquençage peuvent être ré-analysées sans nécessité de reséquençage, en incluant les nouveaux gènes candidats au panel . Cette approche permet donc une réactivité importante et une grande flexibilité. Par exemple, lorsqu'un nouveau gène de prédisposition au cancer est identifié, il suffit d'ajouter informatiquement le gène au panel virtuel pour obtenir très rapidement les variants de ce gène dans la cohorte, y compris de façon rétrospective. De plus, elle permet de gagner en termes d'espace de stockage en ne conservant que les variants des régions cibles.

4- Séquençage du génome complet ou *WGS* pour *Whole Genome Sequencing*

Avec le développement du NGS, des séquenceurs de plus en plus performants arrivent sur le marché. Ainsi, le séquenceur NovaSeq6000 de chez Illumina affiche une capacité de séquençage de près de 3000 Gb en 44 heures, soit l'équivalent de 48 génomes humains. Ces nouvelles machines et le coût diminué de séquençage rendent le séquençage de génomes entiers plus accessible pour les applications diagnostiques.

En tant que technologie de séquençage, le WGS offre un certain nombre d'avantages significatifs par rapport aux autres méthodes, notamment une couverture complète du génome donnant accès aux zones non-codantes du gène comme les introns, les promoteurs et également aux régions régulatrices intergéniques. Cette approche permet également une couverture plus homogène des régions exoniques et la possibilité de détecter des variants structuraux tels que les translocations ou les inversions. En revanche, ce séquençage WGS n'est actuellement pas accessible en laboratoire de diagnostic parce qu'il reste très coûteux et génère un important volume de données ce qui pose des problèmes techniques de stockage et une difficulté d'interprétation du grand nombre de variants générés par génome (3,5 millions de variants nucléotidiques par génome). Il est fort à parier que dans un futur proche, les limites de cette technologie seront au fur et à mesure surmontées par les évolutions technologiques et l'organisation en réseaux nationaux des plateformes de séquençage et d'analyse. Ainsi, avec une réduction des coûts et une rationalisation des analyses de données en aval, le recours au WGS pour la gestion de routine des patients sera possible.

Chapitre III

Interprétation des variants dans le cadre des formes mendéliennes de cancer

I- Classification des variants

Avec le développement et l'implémentation du séquençage à haut-débit de l'ADN à la fois en recherche et en génétique médicale, l'intégralité du patrimoine génétique et en particulier l'exome est devenue accessible. Ainsi, la détection des variants au sein des gènes associés à des maladies mendéliennes est aujourd'hui grandement facilitée malgré la complexité des processus nécessaires au traitement des données. En effet plus le coût du séquençage diminue, plus les données de séquence s'accumulent et avec elles, le nombre de variants découverts. Ces variants peuvent correspondre à des variants délétères, responsables de la maladie mais peuvent également correspondre à de simples polymorphismes. En conséquence, le défi de la génétique moderne n'est donc plus la détection des variations génétiques chez les patients mais leur interprétation biologique et la détermination de leur classe de pathogénicité pour le diagnostic.

Dans ce contexte, le Collège américain de génétique médicale et de génomique (ACMG, American College of Medical Genetics) a réuni en 2013 un groupe de travail composé de représentants de l'ACMG et de l'AMP (Association for Molecular Pathology) pour élaborer des lignes directrices pour l'interprétation des variants de séquence dans le but d'homogénéiser les pratiques au niveau international. Aujourd'hui ces recommandations ont été adaptées par l'ANPGM (Association Nationale de Patriciens de Génétique Moléculaire) et sont appliquées dans les laboratoires de diagnostic génétique en France afin de limiter la variabilité inter- et intra-laboratoire dans l'interprétation des variants.

Selon la dernière version des recommandations révisée en 2015, (Richards et al., 2015; Amendola et al., 2016) et en se basant sur un faisceau d'arguments à prendre en compte afin de classer les variants, le groupe de travail a proposé un système de classification des variants à cinq niveaux relatifs pour les maladies mendéliennes en utilisant des termes spécifiques, allant de « pathogène » à « bénin » afin de les interpréter selon le niveau de leur pathogénicité (**Figure 12**):

1. **Variants bénins ou classe 1**
2. **Variants probablement bénins ou classe 2**
3. **Variants de signification indéterminée (VSI) ou classe 3**
4. **Variants probablement pathogènes ou classes 5**
5. **Variants pathogènes ou classes 4**

Les variants de classes 1 et 2 entraînent peu ou pas de modifications sur la protéine. Il s'agit de polymorphismes bénins (fréquence > 1 % dans la population générale) qui sont communs dans une population en bonne santé.

Les variants de classe 3 ne peuvent pas être interprétés dans l'état actuel des connaissances. La littérature, les bases de données et les outils de prédiction n'ont pas permis de les classer ou les prédictions sont contradictoires. Il n'est pas possible de les relier à la pathologie ni d'exclure leur implication. Ainsi, ces variants ne pourront pas être utiles pour le conseil génétique des patients et de leur famille. Néanmoins, ils peuvent faire l'objet d'études complémentaires visant à apporter des arguments permettant de les reclasser en variant d'une autre classe.

Les variations de classes 4 et 5 entraînent des conséquences sur la protéine et sont considérées comme étant responsables des maladies. Ces variants peuvent concerner des mutations non-sens entraînant l'apparition d'un codon stop prématuré, les insertions ou délétions qui entraînent un décalage du cadre de lecture et les remaniements de grande taille ainsi que les variants décrits pathogènes dans la littérature et les bases de données.

En fonction de l'évolution des connaissances, des variants peuvent être repris et ré-examinés afin de les reclasser dans une nouvelle classe.



Figure 12 : Classement des variants détectés selon les recommandations de l'ACMG

Les variants détectés sont généralement classés sur une échelle de cinq points pour indiquer la probabilité que le variant en question soit associé à une maladie.

Ainsi, cette classification des variations de séquence en différentes catégories facilite l'interprétation de la signification clinique d'un variant.

II- Stratégies d'interprétation des variants

L'évaluation de la pertinence clinique des variants peut être extrêmement difficile. Les mutations **perte de fonction** telles que les mutations non-sens, les mutations décalant le cadre de lecture ou les grands réarrangements géniques ne posent en général pas de problème d'interprétation dans les gènes clairement impliqués dans une pathologie. Les mutations synonymes sont quant à elles généralement classées en variants de classe 1 sauf dans le cas où un impact sur l'épissage est soupçonné (mutations des dernières bases d'exon par exemple). La principale difficulté de l'interprétation des variants réside dans l'analyse des mutations faux-sens. En effet, si ces mutations peuvent correspondre à de simples polymorphismes sans impact sur la fonction de la protéine, elles peuvent également correspondre à des mutations causales à fort effet biologique. Pour classer ces mutations faux-sens plusieurs lignes d'arguments peuvent être apportées pour ou contre l'implication causale du variant.

1) Données épidémiologiques

La première étape consiste à vérifier si ce variant est référencé dans les bases de données généralement accessibles en ligne ou décrit dans la littérature publiée, l'identification préalable de ce variant (si un allèle a déjà été rapporté avec une association de maladie) et de sa fréquence estimée dans la population générale permettant un accès et un partage rapide des connaissances disponibles pour l'interprétation des données de séquençage.

- Bases de données de populations générales

L'avènement du NGS et son accessibilité ont permis le séquençage du génome ou de l'exome de plus en plus d'individus au fur et à mesure des années. Ceci a permis de constituer des bases de données contenant un nombre croissant de variants qui sont continuellement découverts dans le génome humain permettant ainsi leur meilleure caractérisation. Ces différentes bases de données sont fréquemment utilisées pour filtrer les variants selon leur fréquence dans la population générale. En effet, une fréquence allélique élevée d'un variant est en défaveur de son implication dans la maladie et de sa classification comme pathogène.

- La première grande base de données recensant les variants génomiques provenant de diverses espèces fut dbSNP (Single Nucleotide Polymorphisme Database ; <http://www.ncbi.nlm.nih.gov/projects/SNP>), crée en 1998 rassemble les premières informations de

fréquence en population des variants identifiés. Néanmoins, la procédure de soumission trop permissive a conduit à des informations incomplètes ou inexactes. Aujourd'hui cette base référence également les variations selon leur degré de pathogénicité.

- Plusieurs projets de séquençage à grande échelle ont été menés et ont permis la constitution des bases de données tel que le projet 1000Genomes qui rassemble les données sur le séquençage du génome de 2 500 individus d'origines diverses provenant de 26 populations différentes (Sudmant et al., 2015) (<http://www.1000genomes.org/>).

- La base ESP (Exome Sequencing Project, <http://evs.gs.washington.edu/EVS/>) qui comprend actuellement les exomes de 6515 individus américains classés en populations, propose également une mise à disposition de ses données, avec la fréquence en population (et en sous-population d'origine africaine ou européenne) des variants identifiés. Toutefois, cette base comprend en plus des individus sains des patients atteints de diverses pathologies notamment cardio-vasculaires et pulmonaires.

- Le consortium ExAC (Exome Aggregation Consortium ; <http://exac.broadinstitute.org/>) rassemble des données de séquence de 60 706 individus, soit 14 cohortes de patients et 6 groupes de populations, issus de larges projets d'exomes indépendants (populations malades et saines) et accessibles à la communauté scientifique (Lek et al., 2016). Cette base est principalement adaptée à l'interprétation des variants exoniques.

- Récemment ce projet s'est ensuite développé pour former une nouvelle base de données gnomAD (Genome Aggregation Database ; <http://gnomad.broadinstitute.org>) rassemblant les données d'exomes (ExAC) mais aussi de génome (1000Genomes) et qui regroupe aujourd'hui les résultats de 123 136 exomes et 15 496 génomes. Néanmoins, cette base contient des données d'individus ayant développés des cancers, des maladies psychiatriques, des démences ou des maladies inflammatoires (**Figure 13**).

- Une initiative française a initié grâce aux fonds alloués par France Génomique (AAP 2013 : « *The French Exome Project* ») l'établissement de la base de données française FREX (*French Exome Project*). Cette base rassemble les données d'exomes de 600 individus contrôles issus de six régions de France.

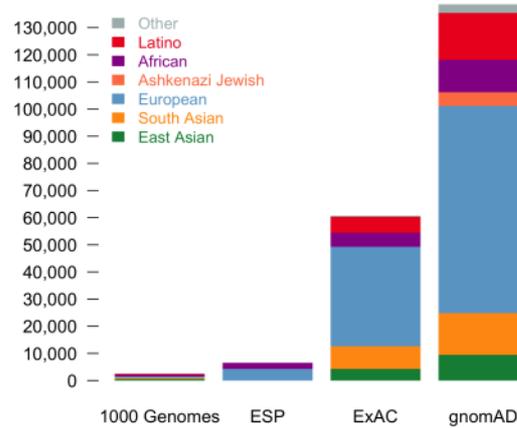


Figure 13 : Taille des bases de données de population générale en nombre d'exomes analysés et répartition des individus selon leur origine ethnique

La base de données la plus récente *gnomAD* dépasse la taille de l'ensemble des autres bases de données précédemment réalisées (1000Genomes, ESP, ExAC).

Une comparaison de la fréquence allélique du variant sera effectuée entre ces populations générales et la population des malades. Une fréquence allélique d'un variant supérieure à 5 % sera un argument en faveur de son caractère bénin. Notons bien qu'il faut bien prendre en compte le mode de transmission de la maladie étudiée vu que pour certaines maladies à transmission autosomique récessive, la fréquence allélique de certains variants pathogènes fréquents pourrait être supérieure à 1 %.

Toutes les informations issues de ces bases de données doivent être utilisées avec prudence car les individus séquencés ne présentent pas de pathologie sévère mais peuvent être atteints de diverses pathologies de l'adulte comme les maladies psychiatriques ou les démences. Il est donc probable que ces bases contiennent des variants pathogènes et pas exclusivement des variants bénins. De plus ces bases de données sont internationales et ne sont donc pas représentatives des variants spécifiques de sous-populations. D'où l'importance de prendre en compte l'origine ethnique du patient pour interpréter les données des bases de données.

- Bases de données des patients

Il existe des bases de données spécifiques de pathologies et issues de réseaux nationaux compilant des informations d'interprétation sur des gènes ou des variants de séquences dans des gènes connus pour être associés à une pathologie. Ces bases offrent ainsi une aide précieuse dans une utilisation médicale. L'avantage de telles bases de données est leur fiabilité grâce à la validation de variants par des groupes d'expert, cependant elles comportent des cohortes plus restreintes que les bases internationales et sont donc moins exhaustives.

Parmi ces bases de données il existe **ClinVar** (<https://www.ncbi.nlm.nih.gov/clinvar/>) qui compile des soumissions de variants réalisées directement par les laboratoires ou par des groupes d'experts et qui comprend à la fois des variants somatiques et constitutionnels (Landrum et al., 2016), la base de données de mutations génétiques humaines **HGMD** (<http://www.hgmd.org>) qui permet d'opérer des recoupements des mutations constitutionnelles associées aux maladies déjà publiées (Stenson et al., 2017), la base de données spécifique de gènes **LOVD** (Leiden Open Variation Database) (Fokkema et al., 2011) ainsi que **COSMIC** (**Catalogue Of Somatic Mutations In Cancer**) qui est la ressource la plus vaste et la plus complète au monde pour explorer l'impact des mutations somatiques sur le cancer humain (Forbes et al., 2017).

- Bases de données spécifiques de gènes

Certaines institutions gèrent des bases de données spécifiques aux gènes qui sont ouvertes au public et qui sont de bonnes ressources pour vérifier si une mutation / variante a déjà été observée, telles que **IARC** (International Agency For Research on Cancer) base centrée sur TP53, fournie pour chaque variant décrit dans la base des informations de sa pathogénicité, **BRCashare BIC** focalisée sur les gènes BRCA1 et BRCA2, et InSiGHT (International Society for Gastrointestinal Hereditary Tumours) qui se focalise sur tous les gènes associés aux pathologies gastro-intestinales héréditaires.

Ces bases qui constituent aussi une ressource précieuse dans l'interprétation des variants doivent néanmoins être utilisées avec précaution vu que les critères définissant un variant pathogène vont pouvoir varier d'une base à l'autre.

2) Orientation clinique et analyse de co-ségrégation

L'évaluation du caractère pathogène d'un variant peut s'appuyer sur une orientation clinique précise et sur l'étude familiale de ségrégation décrite par Thompson *et al.*, 2003. Dans l'idéal, il est nécessaire de pouvoir réaliser la recherche du variant chez les 2 parents du cas index (analyse de trio). Ce test facile à réaliser peut aider à interpréter ce variant.

Dans le cas d'une présentation sporadique, lorsque le variant qui touche un gène associé au cancer n'est retrouvé chez aucun des deux parents sains, il est dit *de novo*, correspondant à un accident génétique survenu sur l'un des gamètes des parents de l'individu ou au niveau

post-zygotique, ce qui constitue un argument supplémentaire en faveur de son implication dans la maladie.

Dans le cas d'une présentation familiale, il conviendra de vérifier la ségrégation de la mutation avec le phénotype étudié. Cet outil puissant pour la classification des variants mesure directement l'association du variant avec le phénotype dans les familles. L'analyse de la ségrégation d'un variant sur plusieurs générations de la famille va permettre de relier le variant ou pas à la maladie. En effet, un variant trouvé chez les membres affectés de la famille mais pas chez les membres non affectés est une preuve forte de sa pathogénicité.

La présence d'une co-ségrégation entre le variant et le phénotype fournit des preuves solides de sa pathogénicité. C'est une méthode robuste car elle est peu sensible aux incertitudes sur la fréquence des allèles, à la stratification de la population ou aux biais de sélection. En revanche, une grande limite de cette approche est la taille des familles vues en consultation qui sont rarement assez informatives. De plus, l'interprétation des données de ségrégation pour les gènes de pénétrance modérée ou incomplète représente un réel défi. Dans ce cas, une évaluation supplémentaire du variant est requise.

3) Prédiction in silico de l'impact des variants

a- Utilisation des algorithmes de prédiction in silico de l'impact sur la protéine

Lorsque le variant est absent des différentes bases de données ou que son implication dans la pathologie n'est pas clairement définie, il est possible de modéliser son impact potentiel sur la fonction de la protéine. Pour cela, de nombreux logiciels de prédiction ont été développés et sont maintenant disponibles pour apporter une aide à l'interprétation, par une évaluation in silico de l'effet fonctionnel de la mutation sur la protéine.

Certains de ces outils sont basés sur des algorithmes prenant en compte la conservation au cours de l'évolution des acides aminés. En effet cet acide aminé conservé au cours de l'évolution a de grandes chances d'appartenir à un élément fonctionnel important dans la protéine. De plus ces outils prennent en compte les structures physico-chimiques et les propriétés biochimiques entre les acides aminés substitués (la polarité, l'encombrement stérique...). Ainsi la conséquence prédite d'une mutation impliquant un changement entre des acides aminés très différents d'un point de vue physico-chimique sera alors plus délétère que si elle concernait des acides aminés proches.

- ❖ Le logiciel SIFT (Sorting Intolerant From Tolerant ; <http://blocks.fhcrc.org/sift/SIFT.html>) est principalement basé sur **l'homologie de séquence protéique** (Ng & Henikoff, 2003). Il considère que les changements dans des positions bien conservées au cours de l'évolution et vers un acide aminé conférant des propriétés physico-chimiques différentes (chargé, polaire...) sont plus probablement associés à un effet délétère sur la fonction de la protéine. La probabilité pour chaque substitution d'acide aminé est calculée pour avoir un score de prédiction SIFT. Dans le cas où les probabilités calculées sont inférieures à un seuil prédéfini, les variants seront considérés comme délétères. Par exemple, un variant faux-sens devrait avoir un effet néfaste sur la fonction de la protéine, lorsque le score SIFT de l'acide aminé substitué est inférieur au seuil de 0,05.
- ❖ L'algorithme PolyPhen-2 (polymorphisme phénotypique, version actuelle PolyPhen-2 : <http://genetics.bwh.harvard.edu/pph2>) est une méthode de prédiction basée aussi sur les homologies de séquences mais également sur la structure de la protéine telle que son repliement et sa structure tridimensionnelle (Adzhubei et al., 2013). Un variant peut avoir un

effet délétère sur la structure si celui-ci touche des résidus clés du repliement protéique. Parmi ces structures on peut citer par exemple les cœurs hydrophobes ou les ponts disulfures. Un ensemble de paramètres (la surface accessible de l'acide aminé, le facteur beta cristallographique, et les modifications de la surface accessible pour les résidus enfouis) sont pris en compte par un classificateur afin de déterminer la probabilité que le variant soit délétère pour la fonction protéique. Ainsi, les variants sont classés comme bénins (benign), possiblement délétères (possibly damaging), ou probablement délétères (probably damaging).

- ❖ MutationTaster2 (<http://www.mutationtaster.org/>, la dernière version du logiciel Web MutationTaster1), qui prend en compte la conservation (nucléotidique et protéique), est conçue pour prédire les conséquences non seulement des substitutions d'acides aminés, mais aussi des modifications introniques et synonymes, des mutations d'insertion et / ou de délétion (indel) et des variants touchant les sites d'épissage (Schwarz et al., 2014). Il utilise un classificateur qui catégorise les variants en variants causaux ou en polymorphismes permettant une évaluation rapide des variants en intégrant les données de fréquence dans les bases de données de populations générales (1000Genomes) et les données de pathogénicité dans les bases de données de patients (ClinVar).

Ces programmes ont tendance à avoir une spécificité plus élevée que la sensibilité, ce qui signifie qu'ils sont meilleurs pour prédire des variants qui ne sont pas pathogènes que ceux qui sont pathogènes. Ainsi, il est essentiel de les combiner, c'est-à-dire d'interroger plusieurs outils et algorithmes, afin d'obtenir la meilleure sensibilité et la meilleure spécificité possible.

Il existe maintenant des logiciels appelés « intégrés », regroupant les prédictions d'un nombre donné d'autres logiciels de prédiction, et utilisant comme critères les scores de prédiction fonctionnelle et de conservation combinés. Parmi lesquels CADD (Combined Annotation Dependent Depletion) (Kircher et al., 2014), Condel (González-Pérez & López-Bigas, 2011) et KGGSeq (Li et al., 2012).

b- Approches d'aide à l'interprétation de variations génétiques sur l'épissage

La conséquence fonctionnelle de la grande majorité des variants identifiés par séquençage à haut-débit demeure difficilement prédictible. Une fraction de ces variants peut avoir un impact sur l'épissage. En plus des mutations des sites introniques d'épissage, certains variants exoniques peuvent avoir un effet sur l'épissage soit en perturbant directement les sites consensus d'épissage, soit en altérant des éléments exoniques de régulation d'épissage (ESR). Plus largement, tous les types de variations nucléotidiques peuvent affecter l'épissage du pré-ARNm en détruisant ou en créant de nouvelles séquences consensus de site d'épissage.

- Analyse in silico

Des outils bio-informatiques utilisant des algorithmes différents ont été développés pour prédire l'effet de variations sur les sites d'épissage (comme MES (MaxEntScan, (Yeo & Burge, 2004), SSF (Splice Site Finder, Shapiro & Senapathy, 1987), NNSplice (Neural Network Splice, Reese et al., 1997), HSF (Human Splicing Finder, Desmet et al., 2009) ou sur les éléments cis régulateurs d'épissage (comme ESEfinder et les approches de prédictions basées sur les scores ESRseq) (Yeo & Burge, 2004). Jusqu'à récemment, et du fait que les éléments régulateurs d'épissage ne soient pas encore bien caractérisés, seules les méthodes de prédictions bio-informatiques dédiées aux sites d'épissage étaient fiables. Ces méthodes ont été développées, au départ, pour caractériser les exons notamment les jonctions intron-exon. Puis, elles ont été détournées de leur but initial afin d'être appliquées dans la sélection des variations à analyser en priorité pour leur effet sur l'épissage par des analyses expérimentales. Ces logiciels de prédiction calculent un score qui renseigne sur la force du site d'épissage à l'état sauvage et mutant en présence d'un variant donné pour prédire l'effet du variant en comparant les scores obtenus.

Récemment, de nouveaux outils bioinformatiques ont été décrits comme prometteurs pour la prédiction de variations affectant potentiellement des éléments exoniques régulateurs de l'épissage. Il s'agit par exemple de l'outil ESEfinder qui vise à prédire la présence d'éléments ESE (Exonic Splicing Enhancer) reconnus par les protéines SR (Serine-arginine Rich proteins) (Cartegni et al., 2003; Smith et al., 2006). Des travaux récents du laboratoire sur différents exons modèles (l'exon 7 de BRCA2 et l'exon 10 de MLH1) ont montré la fiabilité d'au moins une parmi les méthodes récentes évaluées, notamment celle basée sur des scores Δ ESRseq, pour prédire « des mutations ESR » et qui consiste à calculer la différence du total des scores ESRseq (Δ ESRseq) entre l'allèle sauvage et l'allèle muté, en prenant en

considération les 6 hexamères chevauchant la position de la variation d'intérêt (Soukarieh et al., 2016). Toutefois, l'utilisation de ces outils nécessite une optimisation des seuils utilisés pour la discrimination des mutations avec ou sans effet sur l'épissage afin d'évaluer avec plus de précision la sensibilité et la spécificité de ces nouvelles méthodes axées sur les ESR.

Le choix des méthodes de prédiction des sites d'épissage repose sur le fait qu'elles soient facilement utilisables et que leur fiabilité en termes de prédictions ait été validée dans différentes études. Ces outils de prédictions peuvent être interrogés indépendamment sur des sites internet en libre accès ou simultanément à l'aide d'Alamut (les sites dédiés aux sites d'épissage et ESE finder seulement).

La pratique consiste à prédire l'effet du variant à l'aide de plusieurs logiciels et de s'assurer que les prédictions sont concordantes entre elles. Dans le cas de discordance, une étude d'Houdayer et al., 2012 a montré que MES et SSF-like sont ceux qui ont la meilleure performance pour prédire les mutations d'épissage. Ainsi ces outils de prédiction *in silico* ont une part plus ou moins importante dans les décisions d'interprétation et leur utilisation par une approche combinatoire semblant néanmoins plus efficace.

- Analyses expérimentales

Dans la mesure du possible, les études d'ARN sont le meilleur moyen d'interpréter les conséquences d'une mutation d'épissage. Deux approches expérimentales sont utilisées dans l'étude de l'épissage :

- La première technique est l'étude des transcrits par RT-PCR qui consiste à étudier directement l'expression des ARN et identifier des transcrits anormaux. Cette méthode permet de comparer le profil d'épissage de la région d'intérêt, entourant la variation à tester, à celui d'individus contrôles non porteurs de variations dans le gène analysé. Cette technique peut se faire directement sur un tube PAXgene qui permet de stabiliser l'ARN mais également sur des lignées cellulaires dérivées du patient (Exemple lignées lymphoblastoïdes). L'avantage de cette technique est que l'ARN est analysé directement à partir de matériel biologique du patient porteur de la variation. Cependant, la rare disponibilité du matériel et la faible expression de certains gènes dans l'ARN extrait d'un tissu donné sont les limites majeures de cette méthode. De plus, certains transcrits sont difficiles à identifier, par exemple des transcrits contenant un codon stop prématuré qui sont des cibles potentiels de dégradation par le système de surveillance NMD (Nonsense Mediated Decay) (Hartmann et al., 2008; Baralle et al., 2009). Cette difficulté peut être contournée par la possibilité de traiter les lignées

cellulaires avec la puromycine, un inhibiteur de la traduction qui permet d'inhiber le système NMD.

- La deuxième technique basée sur l'utilisation de **minigènes (test ex vivo)** a été développée dans notre laboratoire (Tournier et al., 2008), permettant d'analyser les conséquences sur l'épissage de variants à partir d'ADN génomique. Dans cette technique les exons sauvages ou comportant la variation, flanqués d'extrémités 5' et 3' de 150 paires de base sont amplifiés par PCR à partir de l'ADN génomique et sont clonés dans le vecteur minigène contenant 2 exons séparés d'un intron. Ensuite les constructions sont transfectées de manière transitoire dans des cellules HeLa. Après 24 heures, les ARN sont extraits et transformés en ADNc par RT-PCR puis le profil d'épissage est analysé sur gel d'agarose et les bandes sont ensuite séquencées (Figure 14).

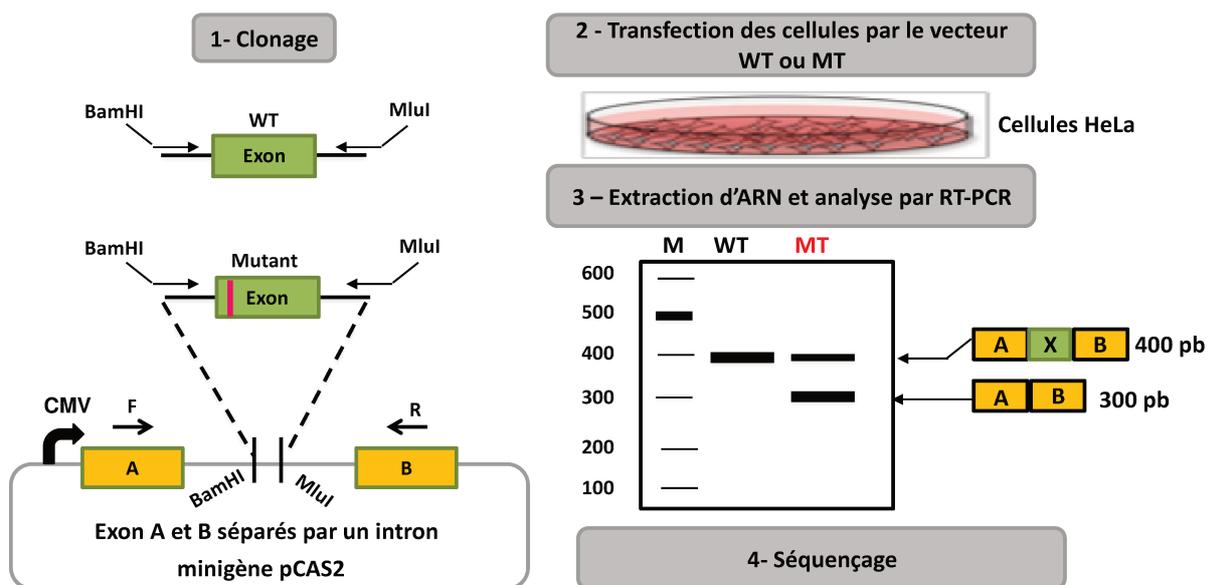


Figure 14 : Les différentes étapes pour la réalisation du test minigène analysant l'effet des variants sur l'épissage

Le fragment d'ADN d'intérêt est amplifié par PCR à partir de l'ADN génomique du patient issu des cellules ou du sang. Ce fragment peut être sauvage (WT, *Wild-type*) ou muté (MT). Les sites de restriction de BamHI et de MluI aux extrémités des amorces de PCR permettent le clonage de cet insert dans l'intron du minigène pCAS2. Le minigène utilisé est constitué de deux exons A et B séparés par un intron. L'expression du minigène est sous le contrôle d'un promoteur de type cytomégalovirus (CMV). Ensuite les vecteurs sont transfectés dans des cellules HeLa, 24h après les ARN sont extraits et l'analyse de RT-PCR est effectuée. Les produits d'amplification sont analysés par électrophorèse sur un gel d'agarose et par séquençage. Si le patient possède un variant responsable d'un défaut d'épissage, il y aura présence de deux bandes correspondant aux deux transcrits obtenus.

Cette approche applicable à tous les gènes présente plusieurs avantages dont l'affranchissement du besoin de l'ARN des patients est le plus important. De plus cette méthode offre la possibilité de tester l'impact d'un variant isolé du reste des variants d'un individu. En revanche les minigènes représentent un modèle artificiel nécessitant un effort considérable vis-à-vis du coût et du temps d'analyse, mais aussi ne prenant pas en compte la nature exacte de ce défaut *in vivo* d'où la nécessité de combiner cette stratégie avec d'autres approches.

4) Co-occurrence avec des variants délétères connus

Dans le cas de maladies génétiques à transmission autosomique dominante la co-occurrence en *trans* d'un variant de signification biologique inconnue avec un variant clairement délétère peut être un argument en défaveur de sa pathogénécité. En effet dans le cas où l'inactivation bi-allélique du gène est connue pour être létale comme c'est le cas pour le gène *BRCA1* (Liu *et al.*, 1996). La coexistence d'un variant délétère et d'un VSI plaident en la faveur du coté bénin du variant. De même, la pathogénécité d'un variant en *trans* dans un même gène, peut être démontrée par la constatation d'un phénotype particulier comme dans le cas du gène *BRCA2* responsable de la maladie de Fanconi (Howlett *et al.*, 2002).

5) Analyses somatiques

Les analyses somatiques sur le tissu tumoral dans le cas de sa disponibilité pourront ajouter un poids déterminant dans l'interprétation des variants constitutionnels. Différentes méthodes ont été développées comme l'analyse de la localisation ou de l'expression des protéines par marquage immunohistochimique spécifique des cellules et des tissus tumoraux, ou la recherche d'instabilité micro-satellitaire dans une région chromosomique du gène d'intérêt comme dans le cas du syndrome de Lynch.

En revanche, il n'existe pas un test fonctionnel simple de routine capable de récapituler toutes les fonctions de la protéine impliquée dans la survenue de la pathologie.

III- Le challenge d'interprétation des variants à l'ère du séquençage tout génome

Malgré les efforts internationaux effectués et la volonté de partager des données cliniques et biologiques par les cliniciens et des généticiens moléculaires pour permettre une reclassification progressive des variants comme variants pathogènes ou variants bénins, l'une des plus importantes limites des laboratoires de génétique médicale effectuant des analyses de gènes impliqués dans les maladies mendéliennes, est l'interprétation d'une grande partie des variants génétiques de signification biologique inconnue. Ces variants sont identifiés avec un nombre plus important que prévu parallèlement à la mise en œuvre de nouvelles technologies de séquençage.

Dans certains cas, l'interprétation de ces VSI peut être fournie par les méthodes déjà citées. Cependant, dans de nombreux cas, l'interprétation reste difficile et ne pourra pas être utilisée pour la prise en charge clinique des patients et des familles. En l'absence d'études fonctionnelles ou d'autres tests de référence, les données disponibles peuvent ne pas être exploitables pour le diagnostic.

Cette interprétation qui impose un lourd fardeau aux professionnels de la génétique du cancer est complexifiée avec la découverte de la variabilité insoupçonnée du génome humain démasquée par le NGS. En effet, chaque exome individuel correspondant à seulement 1,2 % du génome, contient environ 20 000 variations nucléotidiques simples (SNV), après l'élimination des variations polymorphiques déjà répertoriées dans les bases de données, il demeure environ 500 variants rares présents chez moins d'une personne sur mille, et la moitié d'entre eux sont associés à une perte de fonction. En 2016, ces chiffres ont été affinés grâce au consortium ExAC et à la compilation de plus de 60 000 exomes d'individus non apparentés entre eux. Selon ExAC, un individu porte en moyenne 21 451 variations exoniques dont 1 500 avec une fréquence allélique inférieure à 1 %, environ la moitié étant des variations synonymes et l'autre moitié des variations non synonymes regroupant les variants tronquant la protéine et les variants faux-sens. Ainsi la détection d'un variant rare voire privé, même survenu de novo (le taux de variation de novo représente 60 à 80 mutations par génome) est un argument totalement insuffisant pour conclure à sa causalité dans le phénotype observé chez le patient (Howlett et al., 2002).

Ainsi, pour que ce saut technologique à l'ère du NGS soit une véritable révolution médicale au service des patients, il sera essentiel de développer des tests fonctionnels

relativement simples et donc facilement implantables dans une structure hospitalière permettant de fournir un argument supplémentaire pour aider à l'interprétation des VSI. Cela dans le but d'optimiser le conseil génétique et l'adaptation de la prise en charge des patients et de leur famille.

Ce problème est illustré par le diagnostic moléculaire du cancer héréditaire du sein et de l'ovaire, une des formes les plus courantes de cancers héréditaires où les VSI identifiés chez les patients dans les gènes BRCA1/BRCA2 représentent environ 30 % des variations détectées chez les patients et ces gènes sont devenus le paradigme du défi de l'interprétation des VSI en génétique médicale. En plus l'intégration récente du gène TP53 dans les panels de gènes dédiés au cancer du sein va très rapidement augmenter le nombre des variants qui devront être interprétés par la suite.

Chapitre IV

Tests fonctionnels dans le cadre des prédispositions héréditaires

I- Cancer héréditaire du sein et de l'ovaire

I. Aspects moléculaires et cliniques du cancer du sein et de l'ovaire

Le cancer du sein est le cancer le plus fréquent chez la femme devant le cancer colorectal et le cancer du poumon, avec un nombre estimé à 58 968 nouveaux cas en 2017 en France (d'après l'Institut National du Cancer, INCa). On estime aujourd'hui qu'environ une femme sur 9 développera un cancer du sein au cours de sa vie. Cette pathologie est la deuxième cause de décès chez la femme avec 11883 décès estimés en 2017 (18,2 % des décès par cancer chez la femme) (Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017). En revanche, le cancer du sein chez l'homme est extrêmement rare et ne représente que 1 % de la maladie.

Environ 5 à 10 % de la totalité des cancers du sein et de l'ovaire correspondent à une prédisposition héréditaire au cancer connue sous le nom de cancer héréditaire du sein et de l'ovaire (*HBOC, Hereditary Breast and Ovarian Cancer*) (Kobayashi *et al.*, 2013). Ce syndrome est associé non seulement à un risque élevé de cancer du sein et de l'ovaire, mais également à un risque accru d'autres types de cancer comme le cancer du pancréas et de la prostate. Cette prédisposition se transmet d'une manière autosomique dominante (Newman *et al.*, 1988; Claus *et al.*, 1991).

Plusieurs gènes très pénétrants, associés à un risque très élevé de cancer lorsqu'ils sont mutés ont été décrits, dont les premiers découverts dans les années 1990 sont les gènes *BRCA1* et *BRCA2* à l'origine d'environ 10 à 25 % du cancer héréditaire du sein et de l'ovaire (Wooster *et al.*, 1994; Miki *et al.*, 1994). Ces gènes mutés à l'état constitutionnel hétérozygote sont associés à une majoration importante du risque de cancer, essentiellement mammaire et ovarienne chez la femme. En effet, le risque d'atteinte d'un cancer du sein chez les individus porteurs d'une mutation dans les gènes *BRCA* est de 56-84 % en comparaison à 12 % dans la population générale (Smith, 2012). A propos du cancer de l'ovaire, le risque d'atteinte chez les individus porteurs d'une mutation dans les gènes *BRCA1* et *BRCA2* est de 36-46 % et de 10-27 %, respectivement, en comparaison à 1,4 % dans la population générale (Smith, 2012). Chez les hommes de 70 ans, le risque cumulatif estimé de cancer du sein était de 1,2 % pour les porteurs d'une mutation dans *BRCA1* et de 6,8 % pour les porteurs d'une mutation dans *BRCA2* remarquablement plus élevé que celui de la population masculine générale qui est inférieur à 1 % (Tai *et al.*, 2007; Kanakis *et al.*, 2018). De plus, les mutations au niveau des gènes *BRCA* augmentent le risque de cancer de la prostate. En effet,

les hommes porteurs d'une mutation dans *BRCA2* ont 20 % de risque de développer un cancer de la prostate à l'âge de 80 ans et ceux porteurs d'une mutation dans *BRCA1* ont un risque légèrement plus important que celui de la population générale qui est de 16 % (Smith, 2012).

À ce jour, 15 à 20 % des patientes atteintes d'un cancer du sein et/ou de l'ovaire sont porteuses d'une altération génétique constitutionnelle dans les gènes *BRCA*. Le reste des cas pourrait s'expliquer par la présence de mutations dans d'autres gènes. Dans ce contexte, le Groupe Génétique et Cancer a établi un panel de 13 gènes reconnus d'utilité clinique pour les individus à risque héréditaire du cancer du sein et de l'ovaire (Moretta *et al.*, 2018).

Ce panel inclut quatre gènes associés à un risque génétique élevé de cancer du sein en plus des gènes *BRCA* : le gène *PALB2* (*Partner and Localizer of BRCA2*) jouant un rôle de réparation des cassures double brin de l'ADN par recombinaison homologue et dont les mutations confèrent un risque de développer un cancer du sein estimé à 35% à 70 ans (Rahman *et al.*, 2007, p. 2; Lee & Ang, 2014), le gène *TP53* associé au syndrome de Li-Fraumeni intégré récemment dans ce panel compte tenu de son implication dans le développement de cancer du sein chez les femmes jeunes (< 31 ans) avec un taux de mutations de 6 % détecté chez ces femmes (McBride *et al.*, 2014; Bougeard *et al.*, 2015; Mai *et al.*, 2016). Le gène *CDHI* prédisposant aux cancers gastriques présentant un risque élevé de cancer du sein lobulaire avec 42 % à 80 ans (Hansford *et al.*, 2015), et le gène *PTEN* dont les mutations sont retrouvées chez les personnes atteintes du syndrome de Cowden, et qui prédisposent également aux cancers colorectaux et de la thyroïde. Le risque mammaire est élevé en cas de mutation *PTEN* avec 70 % à 70 ans (Bubien *et al.*, 2013) . Les 7 gènes restants sont associés à un risque élevé de cancer de l'ovaire (*RAD51C*, *RAD51D*, *MLH1*, *MSH2*, *MSH6*, *PMS2* et *EPCAM*) (Bonadona *et al.*, 2011; Loveday *et al.*, 2012).

L'analyse de ce panel est recommandée devant tout contexte évocateur d'une prédisposition héréditaire aux cancers du sein ou de l'ovaire :

- Le nombre de cas de cancers du sein chez des apparentés du premier ou du deuxième degré dans la même branche parentale
- La précocité de survenue du cancer du sein (40 ans ou moins)
- La bilatéralité de l'atteinte mammaire (notamment avant 65 ans)
- La présence de cancer(s) de l'ovaire
- L'observation éventuelle de cancer(s) du sein chez l'homme

Les personnes porteuses d'une altération génétique constitutionnelle en lien avec le syndrome seins-ovaires doivent se voir proposer une stratégie de suivi spécifique, basée sur la surveillance et/ou la chirurgie préventive. Aujourd'hui les individus porteurs d'une mutation pathogène dans l'un des gènes de prédisposition peuvent bénéficier de stratégies d'évaluation et de gestion des risques, notamment une surveillance accrue du cancer avec une IRM mammaire annuelle dès 30 ans, une mastectomie prophylactique avant 40 ans ou une annexectomie prophylactique après 40 ans et tout projet parental (D'après l'INCA).

L'identification d'une altération délétère des gènes *BRCA* est d'autant plus cruciale maintenant que des thérapies ciblées peuvent être proposées aux femmes porteuses d'une altération constitutionnelle des gènes *BRCA*. Il s'agit par exemple des inhibiteurs de PARP (l'olaparib), déjà développés dans le chapitre I, et qui sont utilisés pour induire une létalité synthétique dans les tumeurs déficientes en *BRCA*.

Malheureusement, les patients qui répondent initialement au PARPi développent souvent une résistance entraînant une rechute du cancer (Lord & Ashworth, 2013; Sonnenblick *et al.*, 2015). Très récemment, il a été suggéré que les inhibiteurs de RAD52 (RAD52i) additionnés aux PARPi, améliorent le résultat thérapeutique dans le cadre des malignités déficientes en *BRCA* en atténuant les autres systèmes de réparation (comme SSA) déclenchés dans les cellules, tout en provoquant une toxicité minimale pour les cellules et les tissus normaux (Sullivan-Reed *et al.*, 2018) (**Figure 15**).

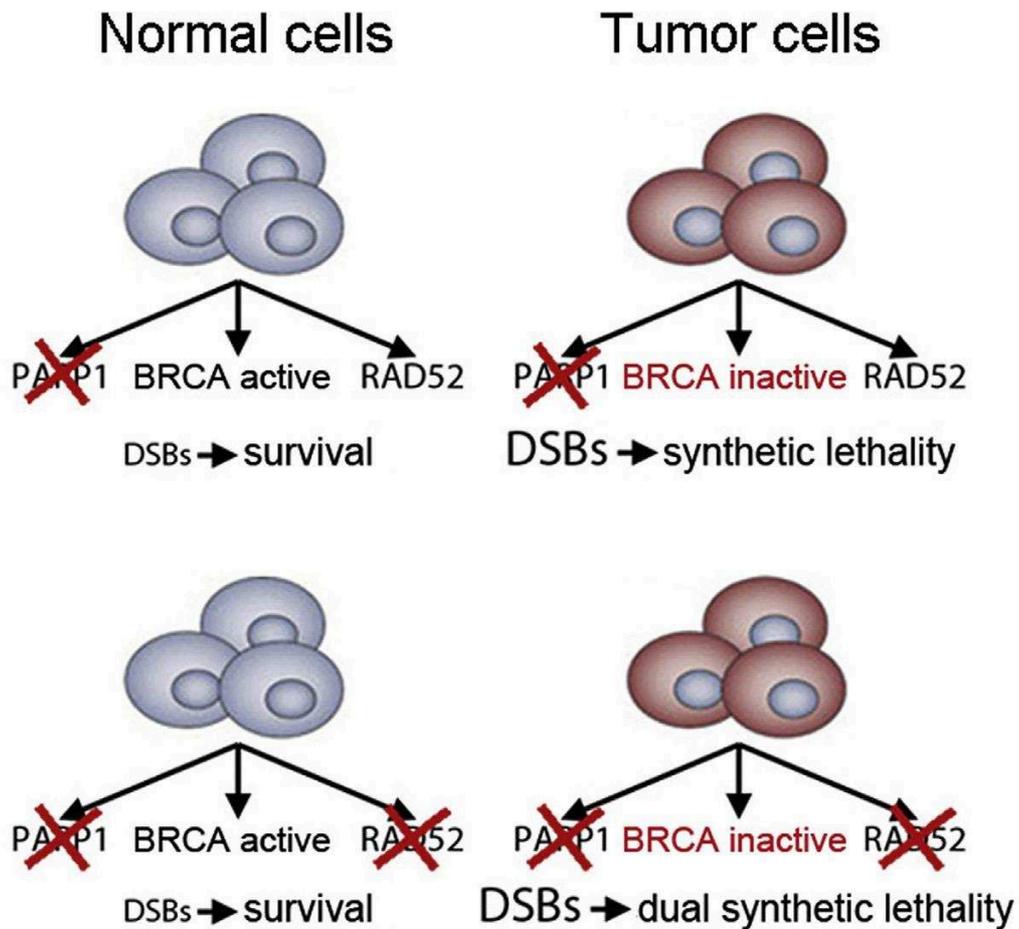


Figure 15 : Double létalité synthétique dans les cellules *BRCA* déficientes (D'après Sullivan-Reed *et al.*, 2018)

Les cellules déficientes en *BRCA* ne pouvaient pas être complètement éradiquées par l'inhibiteur de PARP ou celui de RAD51 seul. Le ciblage simultané de PARP1 et de RAD52 entraîne une accumulation synergique de cassures double brin de l'ADN et exerce une double létalité synthétique provoquant l'éradication de cellules tumorales déficientes en *BRCA*.

II. Structure et fonction des gènes *BRCA*

Les deux gènes suppresseurs de tumeur *BRCA1* et *BRCA2* ont été identifiés pour la première fois en 1994 et 1995 par des études de liaison menées chez des familles de cancers du sein multiples (Miki et al., 1994; Wooster et al., 1995). Ils sont localisés respectivement en 17q21 et 13q12-13, ayant des structures génomiques proches, notamment avec des exons géants. Ils ne présentent cependant aucune homologie de séquence. Ces gènes, exprimés de façon ubiquitaire chez l'homme, codent chacun pour une phosphoprotéine nucléaire jouant un rôle dans le maintien de la stabilité génomique, la régulation de la transcription, le contrôle du cycle cellulaire, le remodelage de la chromatine et un rôle essentiel dans la réparation des cassures double brin de l'ADN par la voie de recombinaison homologe (HR).

a- Le gène *BRCA1*

Le gène humain *BRCA1* s'étend sur environ 80 kb et comporte 23 exons dont 22 codent pour un grand polypeptide de 1863 acides aminés (exon 1 non codant). L'exon 11 représente à lui seul 60 % de sa séquence codante et les autres exons sont de plus petites tailles. Les séquences introniques représentent 90 % du gène et contiennent une forte quantité de séquences répétées de type *Alu* (42 %).

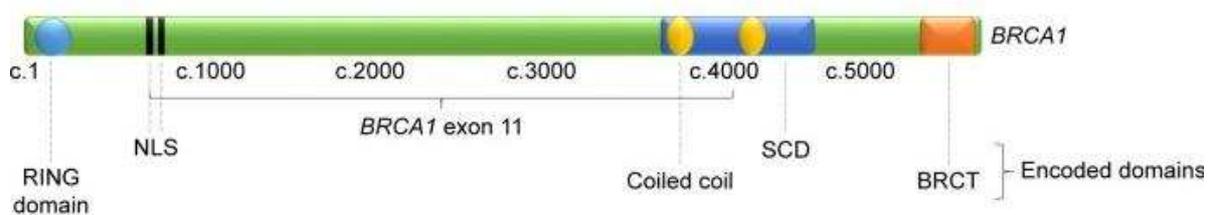


Figure 16 : Structure du gène *BRCA1* et domaines protéiques identifiés de la protéine *BRCA1* (D'après Hollis et al., 2017)

NM_007294.3; NLS: *Nuclear Localization Signal*; BRCT: *BRCA1 C-Terminal Domain*; SCD: *Serine Cluster Domain*.

b- La protéine BRCA1

La protéine BRCA1 fait 190 kDa. Seules ses régions N- et C- terminales sont fortement conservées entre les espèces.

Le domaine *RING finger* en N-terminal (acides aminés 9 à 98) riche en cystéine contient un domaine de liaison au zinc présent dans diverses protéines régulatrices. Ce domaine possède une activité E3 ubiquitine ligase des protéines cibles. Il est indispensable aux interactions protéiques notamment avec la protéine BARD1 (*BRCA1 Associated Ring Domain 1*) qui possède également ce domaine *RING finger*. Cette interaction est essentielle pour la stabilité de ces deux protéines.

Le domaine BRCT (*BRCA1 Carboxy-Terminal domain*) (1649-1734 et 1758-1859) retrouvé à l'extrémité C-terminale est constitué d'une répétition en tandem de deux motifs BRCT pouvant fixer des phosphopeptides et donc permettre des interactions avec des protéines phosphorylées. Ce domaine forme plusieurs complexes différents avec différentes protéines comme le protéine BRIP1, CtIP et Abraxas jouant un rôle dans la résection de l'ADN et le recrutement de la RPA (*Replication Protein A*) sur le site endommagé (Yu *et al.*, 1998; Cantor *et al.*, 2001).

Elle possède également **un site de liaison à l'ADN** situé au centre de la protéine et codé par l'exon 11, **deux signaux de localisation nucléaire NLS** (503-508 et 607- 615) facilitant la localisation de BRCA1 au noyau et **un domaine superhélice ou coiled-coil (CC)** (1393-1424) permettant la fixation de la protéine PALB2, le partenaire et le localisateur de BRCA2 facilitant le recrutement de RAD51 aux cassures double brin de l'ADN (Roy *et al.*, 2011; Prakash *et al.*, 2015; Nepomuceno *et al.*, 2017; Takaoka & Miki, 2018).

c- Le gène BRCA2

Le gène *BRCA2* s'étend quant à lui sur 84 kb et il contient 27 exons, dont 26 sont codants. Il code pour une des plus grandes protéines du protéome avec 3418 acides aminés. Il est lui aussi caractérisé par un exon 11 représentant environ 50 % de la séquence codante, par sa richesse en séquences répétées mais qui est moins élevée en séquence *Alu* que celle de *BRCA1* (20 %).

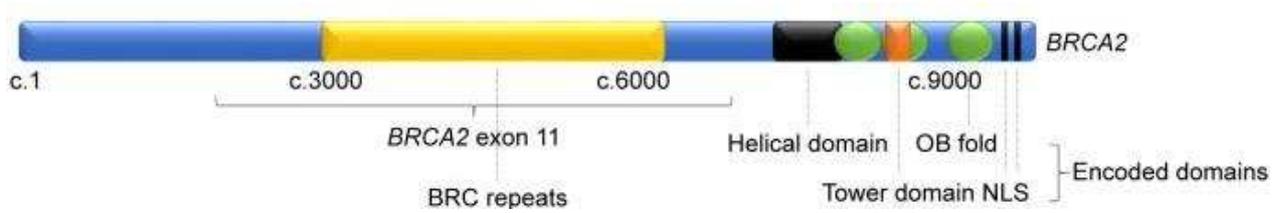


Figure 17 : Structure du gène *BRCA2* et domaines protéiques identifiés de la protéine *BRCA2* (D'après Hollis *et al.*, 2017)

NM_00059.3; NLS: *Nuclear Localization Signal*; OB: *Oligonucleotide Binding*.

d- La protéine BRCA2

L'analyse de la séquence protéique de BRCA2 a permis de mettre en évidence la présence (i) d'un **domaine N-terminal** (résidus 10–40) qui interagit avec la protéine PALB2 essentielle pour l'activité de recombinaison homologue, et la protéine EMSY (*BRCA2 interacting transcriptional repressor*) qui inhibe sa fonction d'activation de la transcription par sa fixation au niveau de cette région, (ii) d'un **motif répété 8 fois** appelé **BRC** d'environ 70 acides aminés dans la région centrale de BRCA2 (résidus 1008-2082) permettant la liaison avec RAD51 (Bork *et al.*, 1996). Cette région abrite également une séquence d'exportation nucléaire NES (résidus 1383–1393) nécessaire à la localisation au centrosome de BRCA2 (Han *et al.*, 2008), (iii) d'un site de fixation à RAD51 dans la portion **C-terminale** (acides aminés 2620-3418) qui comprend une région d'environ 800 acides aminés, appelée **DBD** (*DNA Binding Domain*). Cette partie la plus conservée de BRCA2 est formée de plusieurs domaines : le **domaine hélical**, trois domaines **OB** (*Oligonucleotide Binding*) liant l'ADN simple brin et double brin, deux signaux de localisation nucléaire **NLS** ainsi qu'un site de phosphorylation par des kinases dépendantes des cyclines (CDK; *Cyclin-Dependent Kinases*) qui régule l'association de RAD51 (Yano *et al.*, 2000; Roy *et al.*, 2011; Martinez *et al.*, 2015; Hollis *et al.*, 2017).

e- Fonctions des protéines BRCA1 et BRCA2

Ces deux protéines BRCA1 et BRCA2 exprimées de façon ubiquitaire chez l'être humain sont des suppresseurs de tumeurs agissant en tant que « caretakers » et dont l'inactivation crée un état permissif dans lequel la cellule accumule des défauts cellulaires qui

amènent une instabilité chromosomique. Au niveau cellulaire, les protéines BRCA1 et BRCA2 jouent des fonctions critiques dans l'intégrité du génome. Cet aspect du maintien de l'intégrité génomique est médié par un réseau cellulaire d'évènements de signalisation qui est déclenché en réponse au stress génotoxique. Les cassures double brin de l'ADN impliquent des protéines capables de détecter les extrémités cassées, des effecteurs qui exécutent la réparation et des médiateurs qui facilitent les interactions entre les senseurs et les effecteurs. Dans les cellules de mammifères, les protéines BRCA jouent un rôle essentiel dans la réparation des cassures double brin de l'ADN par la voie de réparation de la recombinaison homologue (HR). La recombinaison homologue est active pendant la phases S et G2 du cycle cellulaire, lorsqu'une chromatide sœur intacte peut servir de modèle pour la réparation. Ce système est également essentiel pour maintenir la fidélité de la réplication. Cette voie dépend de plusieurs suppresseurs dont les gènes *BRCA1* et *BRCA2*, qui facilitent les différentes étapes de la réaction de réparation (Roy *et al.*, 2011, p. 2 ; Sung, 2018).

La réparation des DSB commence par la dégradation ou la résection d'un brin de l'ADN au site de la lésion, créant une étendue d'ADN simple brin. En effet, à la suite d'une lésion de l'ADN, le complexe Mre11-RAD50-Nbs1 (MRN) effectuant la résection se lie aux extrémités des sites DSB et recrute la protéine ATM (*Ataxia Telangiectasia Mutated*) capteuse des lésions à l'ADN. ATM phosphorylera plusieurs cibles et principalement l'histone H2AX en réponse aux lésions de l'ADN. L'histone H2AX phosphorylée, également appelée γ -H2AX, recrute BRCA1 et RAD51, pour former des foyers nucléaires sur les sites de lésions de l'ADN. BRCA1 va ensuite interagir avec le complexe CtIP et le complexe MRN favorisant ainsi la catalyse de la résection des extrémités de l'ADN. BRCA1 contribue également au choix de la voie de réparation via la modulation de la résection terminale en empêchant 53BP1 d'accéder aux extrémités de l'ADN et de diriger la réparation vers le NHEJ (*Non Homologous End Joining*), système de réparation moins fiable des cassures double brin. Une fois la résection terminée, le complexe RPA est recruté pour favoriser la formation de nucléofilaments en éliminant la formation de structures secondaires et en protégeant les extrémités de l'ADN de la dégradation. Ensuite, certains médiateurs régulent positivement le déplacement RPA, afin de faciliter le recrutement de RAD51. BRCA2 est notamment étroitement impliquée dans cette étape par le biais de son interaction via ses domaines BRC avec la recombinase RAD51. BRCA2 serait nécessaire à la localisation et au fonctionnement adéquat de RAD51 au niveau des foyers de lésions de l'ADN (Martinez-Useros & Garcia-Foncillas, 2016). Les molécules RAD51 se lient alors à l'ADN simple brin pour former des

structures filamenteuses. Ces foyers RAD51 favorisent la reconnaissance des séquences homologues sur la chromatide sœur, et catalysent l'invasion du duplex qui servira de matrice (Chen *et al.*, 2018; Sharma *et al.*, 2018) (**Figure 18**).

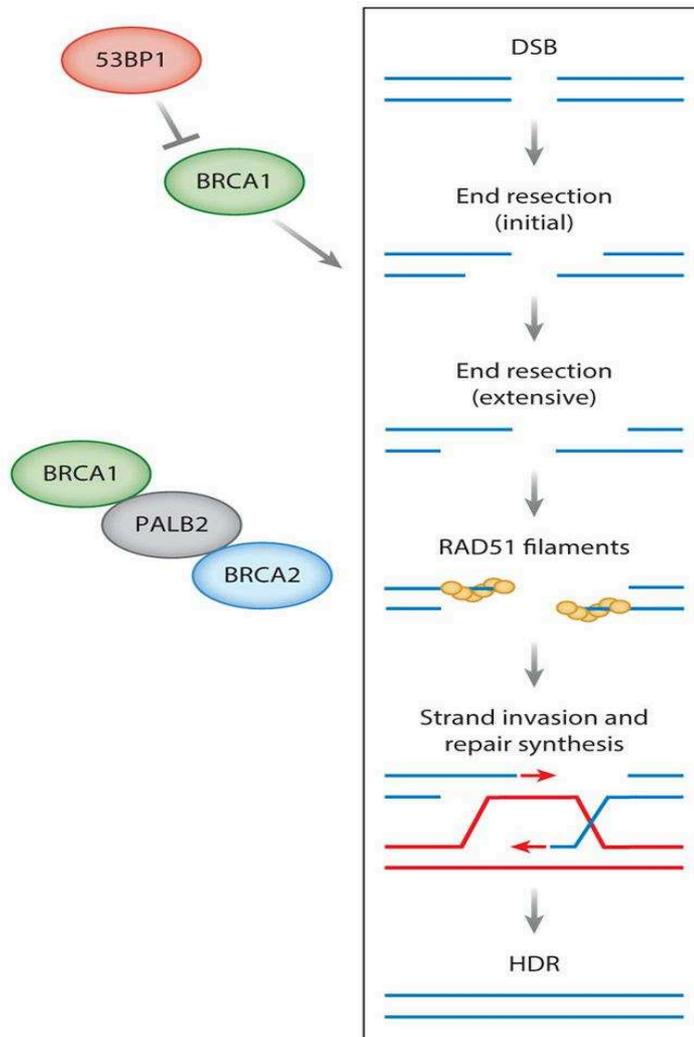


Figure 18 : La réparation des cassures double brin de l'ADN par recombinaison homologue (HR) (D'après Chen *et al.*, 2018)

À la suite des cassures double brin (DSB) de l'ADN, un déterminant précoce du choix de la voie de réparation des DSB est la résection des extrémités de l'ADN. Une résection des extrémités est régie par BRCA1 en concurrence avec la protéine antagoniste 53BP1 afin de favoriser l'engagement de la voie de réparation par recombinaison homologue (HR, également appelée HDR), voie majeure de réparation des DSB. La résection finale peut avoir lieu en deux étapes, une phase initiale et une phase plus extensive entraînant des étendues d'ADN simple brin. BRCA1 favorise également les étapes ultérieures du HR via son interaction avec la protéine PALB2, qui à son tour recrute BRCA2. Ensuite, BRCA2 va contrôler la formation de filaments RAD51 sur le brin simple d'ADN, suivie de l'invasion du brin typiquement sur la chromatide sœur, qui servira de matrice, pour la synthèse et la réparation de l'ADN.

Les protéines BRCA sont aussi impliquées dans la voie de l'anémie de Fanconi, voie de réparation des lésions entre les brins ICL (*Interstrand CrossLinks*) les plus délétères, et des lésions toxiques qui empêchent la réplication et la transcription en inhibant la séparation des brins de l'ADN (Schlacher *et al.*, 2012; Rodríguez & D'Andrea, 2017). L'altération de cette voie entraîne une anémie de Fanconi (FA) qui est un syndrome héréditaire autosomique récessif caractérisé par des anomalies congénitales et du développement, une insuffisance de la moelle osseuse et une forte prédisposition au cancer (Deans & West, 2011). Au cours de cette voie de réparation, les ICL sont d'abord traités par une incision locale favorisée par l'ubiquitination de FANCD2, ce qui génère une cassure double brin. La voie de recombinaison homologue est nécessaire pour la réparation de ces DSB. Pour favoriser la réparation fidèle, les protéines FA telles que FANCD2 et BRCA1 favorisent la HR en inhibant la liaison aux extrémités des DSB des facteurs de la voie de réparation sujette à la voie de recombinaison non homologue (NHEJ) (Adamo *et al.*, 2010). De plus, les protéines BRCA2 (FANCD1) et BRCA1 (FANCS) sont essentielles dans cette voie de réparation en favorisant le chargement de RAD51 (FANCR) sur l'ADN réséqué protégeant ainsi les brins naissants de la dégradation au niveau des fourches bloquées. En l'absence de ces protéines, les brins naissants peuvent devenir un substrat de dégradation par MRE11 et d'autres nucléases, entraînant une instabilité du génome (Kim & D'Andrea, 2012; Schlacher *et al.*, 2012).

En plus du rôle de BRCA1 dans la réparation des cassures double brin, elle intervient dans plusieurs mécanismes cellulaires essentiels pour la stabilité du génome en interagissant avec un grand nombre de protéines. Grâce à son domaine RING finger et à son interaction avec BARD1, ce complexe hétérodimère exerce une activité d'ubiquitine ligase sur plusieurs substrats dont l'histone H2A, favorisant l'étape de résection dans la recombinaison homologue et la réparation de l'ADN (Densham *et al.*, 2016, p. 1). BRCA1 intervient aussi dans la régulation de la transcription de nombreux gènes cibles grâce à son domaine BRCT (Monteiro *et al.*, 1996; Monteiro, 2000) et dans le remodelage de la chromatine en jouant un rôle dans la régulation de la condensation de cette dernière (Zhu *et al.*, 2011). BRCA1 est impliquée dans la régulation du cycle cellulaire dans toutes les phases du cycle et notamment en formant le complexe BRCA1 – BACH1 requis pour les points de contrôle de la réplication pendant la phase S en réponse à des fourches de réplication bloquées avec TOPBP1 (Kumaraswamy & Shiekhattar, 2007; Gong *et al.*, 2010, p. 1). Elle participe au maintien du chromosome X inactif par le recrutement du transcrit XIST (X Inactive Specific Transcript) (Silver *et al.*, 2007), au processus de duplication du centrosome qui est un évènement

responsable de la ségrégation égale des chromosomes lors de la division nucléaire d'une cellule (Starita et al., 2004). BRCA1 est également impliquée dans la voie de l'apoptose en activant les séquences régulatrices d'apoptose GADD45 (Harkin et al., 1999). Une étude récente a montré que le récepteur du peptide natriurétique 3 (NPR3) protège les cardiomyocytes de l'apoptose en inhibant BRCA1 et TNF- α cytosoliques (Lin et al., 2016). Récemment, une fonction entièrement nouvelle de BRCA1 impliquée dans la mitophagie a été mise en évidence (Sumpter et al., 2016). La mitophagie est une autophagie sélective, spécifique des mitochondries, éliminant les mitochondries dépolarisées et endommagées dans le but de limiter la génération excessive d'espèces réactives de l'oxygène ROS (Reactive Oxygen Species) (Bialik et al., 2018). Des explorations plus détaillées de cette nouvelle fonction de BRCA1 lors de la maintenance de l'homéostasie cellulaire pourraient fournir de nouvelles orientations pour l'étude de la physiopathologie des cancers résultant de mutations du gène BRCA1, pour lesquelles la signification clinique n'a pas été clarifiée.

La protéine BRCA2 qui forme des complexes de poids moléculaire élevé avec d'autres protéines confère de nombreuses fonctions (Martinez et al., 2015; Shailani et al., 2018). Elle intervient dans plusieurs mécanismes cellulaires comme la régulation de la transcription (Milner et al., 1997). Par exemple, BRCA2 coopère avec BRCA1 et l'histone acétyltransférase P / CAF (P300/CBP-Associated Factor) pour améliorer la transactivation induite par les récepteurs d'androgènes (AR) et GRIP1 (Glutamate Receptor Interacting Protein 1), coactivateur de l'AR, favorisant ainsi l'effet anti-prolifératif de l'AR (Shin & Verma, 2003). En se liant par son extrémité N-terminale avec P/CAF, composant essentiel du point de contrôle SAC (Spindle Assembly Checkpoint) en mitose, BRCA2 est impliquée dans la maintenance du point de contrôle G2/M en cas des lésions de l'ADN (Lin et al., 2003). BRCA2 est aussi impliquée dans le traitement ou la prévention des hybrides ARN-ADN qui sont des sous-produits normaux de la transcription mais dont leur persistance pourra être à l'origine du blocage de la réplication. En effet, BRCA2 interagit avec les protéines DSC1 et PCID2 et d'autres facteurs, constituant le complexe TREX-2. Ce dernier participe à la biogenèse et à l'exportation de ribonucléoprotéines messagères (mRNP) (Bhatia et al., 2014). Elle participe également à la cytokinèse (Daniels et al., 2004). Il a été montré que BRCA2 phosphorylée par PLK1 (Polo-Like Kinase 1) permet son interaction avec la Myosine IIC (NMCII), une protéine motrice moléculaire qui se lie à l'actine cytosquelettique et régule la cytokinèse (Takaoka et al., 2014). Cela explique l'observation d'une instabilité

chromosomique avec une incidence élevée de cellules binucléées dans les cellules déficientes en BRCA2 (Lee et al., 2011).

III. Tests fonctionnels spécifiques des gènes *BRCA*

Les gènes BRCA1 et BRCA2 présentent un large spectre mutationnel. Il existe plusieurs bases de données qui recensent les variants génétiques constitutionnels dans BRCA1 et BRCA2 détectés chez les patientes atteintes d'un cancer du sein ou de l'ovaire. Il y a actuellement **11793** variations constitutionnelles distinctes de BRCA1 énumérées dans la base de données française BRCA Share (<http://umd.be/BRCA1/>), auparavant UMD BRCA1/BRCA2 (Bérout et al., 2016). Parmi ceux-ci, **2414 (20 %)** sont des variants de signification biologique inconnue (VSI) qui ne peuvent pas être utilisés pour le conseil génétique. Quant à BRCA2, parmi les **15955** variations constitutionnelles répertoriées dans la base de données Française, **4577 (28 %)** sont des VSI (<http://umd.be/BRCA2/>).

Ces VSI correspondent en majorité à des variants faux-sens, mais comprennent également des variants pouvant affecter les taux d'ARN messager pleine longueur avec des effets plus ou moins partiels. De plus, le plus souvent, les analyses de co-ségrégation qui permettraient de trancher quant à la pathogénicité de ces variants ne peuvent pas être menées à cause de la petite taille des familles vues en consultation. Dans ce cas, la collecte des informations sur d'autres familles portant le variant peut être très informative. Cependant, la plupart des VSI détectés dans les gènes BRCA sont des variants rares et sont donc peu détectés chez plusieurs familles. Cette interprétation est encore complexifiée par la pénétrance incomplète de la pathologie. Le besoin critique d'approches alternatives pour classer les VSI a stimulé plusieurs groupes de recherche à développer des tests in vitro pouvant être utilisés pour tester l'impact fonctionnel des variants détectés dans les gènes BRCA1 et BRCA2 (Farrugia et al., 2008; Millot et al., 2012; Guidugli et al., 2014; Toland & Andreassen, 2017). Le challenge de ces tests fonctionnels réside dans la multiplicité des fonctions connues des protéines BRCA1 et BRCA2. La plupart de ces fonctions de BRCA1 et BRCA2 décrites ci-dessus ont servi de base aux analyses fonctionnelles des variants de ces gènes.

A- Test de réparation directe par recombinaison homologue (Homology-Directed Repair, HDR)

Ces tests basés sur l'activité clé des protéines BRCA1 et BRCA2 de réparation des cassures double brin de l'ADN par la recombinaison homologue sont les plus couramment utilisés. Ils reposent sur la présence de 2 allèles inactifs encodant le gène de la protéine fluorescente verte (GFP, Green Fluorescent Protein) dans le génome de cellules rendues déficientes en BRCA1 ou BRCA2 par la transfection d'un ARN interférant (RNAsi small interfering). Un plasmide contenant la séquence codante du gène BRCA1 ou BRCA2 portant la variation à tester est transfecté dans les cellules, puis une cassure double brin est induite au niveau du cDNA de la GFP par l'utilisation d'une endonucléase (I-Sce1) exprimée par un plasmide. Si les protéines BRCA1 et BRCA2 sont fonctionnelles, la recombinaison homologue sera efficace et l'allèle de la GFP clivée sera réparé à l'aide de la deuxième copie. Si le variant testé impacte de façon délétère la fonction de BRCA1 ou BRCA2 sur la recombinaison homologue, la GFP ne sera pas réparée et les cellules n'exprimeront pas la GFP (**Figure 19**).

Cet essai a été développé dans des lignées cellulaires de fibroblastes de poumon d'hamster V-C8 (BRCA2-déficient VC8 cells) (Wu et al., 2005b; Farrugia et al., 2008; Guidugli et al., 2013) et également sur des cellules humaines (HeLa-DR13-9 cells) (Ransburgh et al., 2010; Starita et al., 2018).

Ces tests représentent un outil très intéressant pour apporter des arguments fonctionnels en faveur ou en défaveur de la pathogénicité d'un variant. Ils sont cependant basés sur l'utilisation de l'ADN complémentaire du gène et ne permettent donc pas de tester tous les VSI intervenant dans les régions non codantes de ces gènes et ne permettent pas d'évaluer l'impact de variations exoniques sur l'épissage. De plus, la transfection ne permet pas de maîtriser efficacement les niveaux d'expression des protéines nulles ni de refléter la présence d'une éventuelle haploinsuffisance.

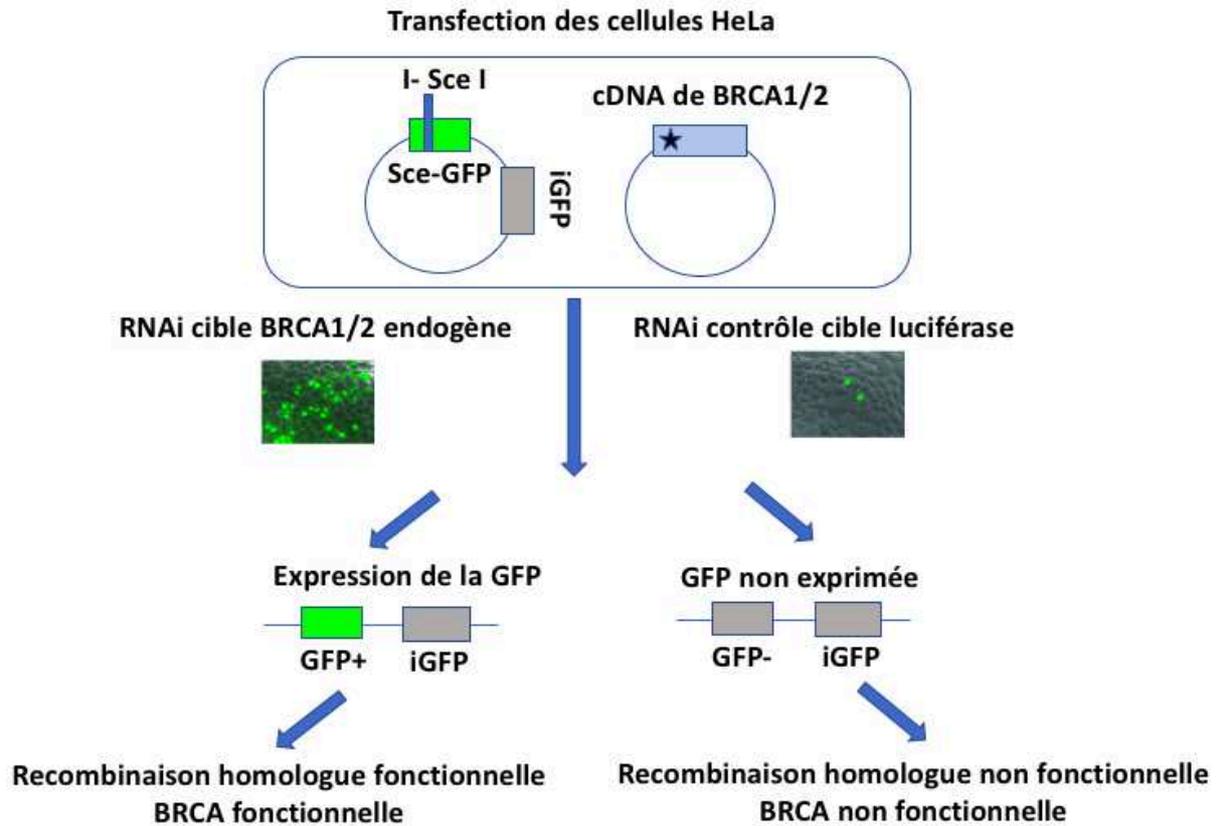


Figure 19 : Principe de l'essai HDR

Dans ce test HDR, le cDNA de BRCA1 ou BRCA2 portant la mutation à tester ainsi que deux allèles inactifs de GFP sont tous transfectés dans le génome des cellules humaines HeLa. La protéine BRCA1 endogène a été réprimée par un ARN interférant ciblant le 3'-UTR de l'ARN et comme contrôle négatif, les cellules sont transfectées par un RNAi ciblant la luciférase. Ensuite, les cellules sont transfectées par un plasmide exprimant I-SceI qui entraîne une cassure double brin au niveau d'un des deux allèles de la GFP contenant un élément de réponse à l'endonucléase. Cette rupture peut être réparée par recombinaison homologue en utilisant le deuxième allèle inactif de GFP. La protéine GFP active est ainsi détectée par fluorescence lorsque la réparation par recombinaison homologue est efficace.

De façon remarquable, une version multiplexe à haut-débit de cet essai a été très récemment publiée (Starita et al., 2018). La différence dans cet essai qui est basé sur le même principe consiste à utiliser la lignée cellulaire rapportrice HeLa HDR conférant un système Flip-recombinase (HeLa-DR-FRT), capable d'intégrer un seul variant à un site commun dans chaque cellule. Des centaines de variants au cours d'une seule expérience pourront être testés par la transfection d'un pool de plasmides contenant les variants BRCA1. Ensuite une étape de sélection des cellules contenant un variant intégré de BRCA1, est effectuée suivie par une analyse des scores d'expression de la GFP (**Figure 20**). Grâce à ce test, l'effet de 1056 substitutions d'acides aminés des résidus 2 à 192 de BRCA1 sur l'activité de réparation par recombinaison homologue a pu être évalué.

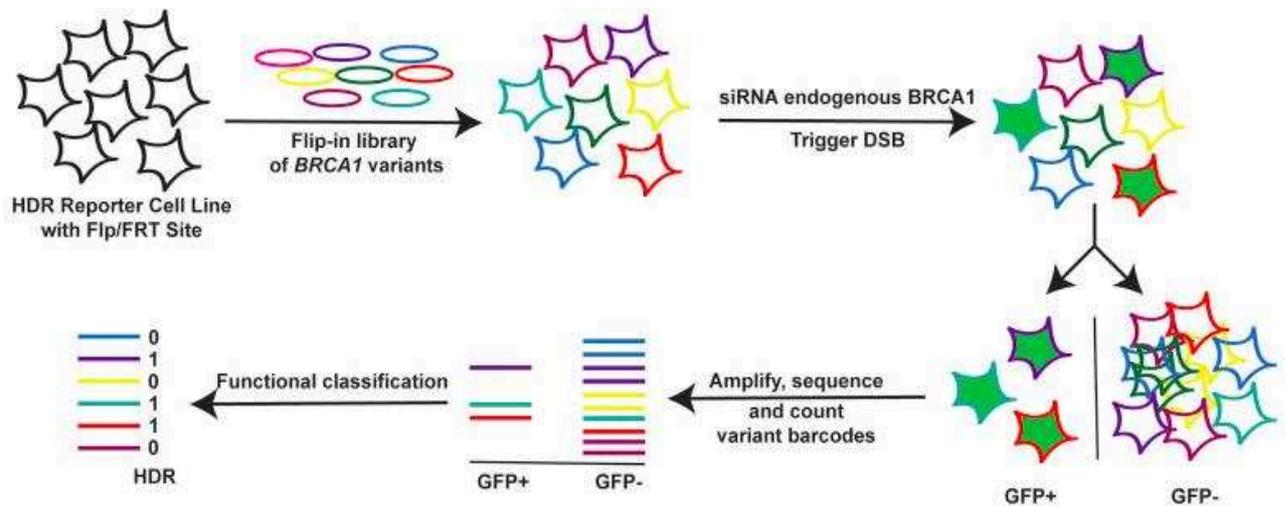


Figure 20 : Principe du test multiplexe de HDR (D'après Starita *et al.*, 2018)

Les variants de BRCA1 sont intégrés dans le génome de la lignée cellulaire HeLa qui confère un système Flip-recombinase. Ces cellules sont dépendantes du variant de BRCA1 intégré après transfection d'ARNsi ciblant l'allèle endogène de BRCA1. Après induction des cassures double brin, les cellules sont triées. Les codes-barres marquant les variants de chaque population sont amplifiés et séquencés. Ensuite, une analyse des scores d'expression de la GFP est effectuée afin de classer les variants de BRCA1 testés.

De façon parallèle à ces essais, des tests basés sur des lignées lymphoblastoïdes issues de patientes porteuses d'une mutation dans l'un des gènes *BRCA* ont été développés pour tester l'impact des variants dans le contexte génomique des patients en s'inspirant de l'implication des protéines BRCA dans la réparation des cassures double brin (Vaclová *et al.*, 2015).

Le principe de ce test consiste à induire dans ces cellules des cassures double brin de l'ADN par des radiations ionisantes réparées principalement par la voie de recombinaison homologue. À la suite de ces lésions, dans le contexte sauvage, le complexe MRN (*Mre11*, *Rad50*, *Nbs1*) se lie aux extrémités des sites des cassures double brin et recrute le capteur de lésions de l'ADN, l'ataxie télangiectasie mutée (ATM). Le rôle clé de l'ATM consiste à recruter des protéines de réparation dans les foyers de lésions de l'ADN, en phosphorylant H2AX, un membre de la famille des histones H2A. Cette phosphorylation se produit au niveau de la sérine 139 et de la thréonine 137 de H2AX dans les quelques minutes qui suivent l'endommagement de l'ADN. Le degré de phosphorylation de H2AX est positivement corrélé à la gradation des lésions de l'ADN (**Figure 21**). H2AX phosphorylée, également appelée γ -H2AX, recrutera des protéines de réparation, telles que RAD51, qui est une protéine cruciale impliquée en aval dans la réparation, et relocalisée dans le noyau en réponse aux lésions de l'ADN pour former des foyers induits par les rayonnements ionisants RAD51 (IRIF, *IRradiation Induced Foci*) visualisés par microscopie immunofluorescente (**Figure 22**). Les cellules déficientes en BRCA ne forment pas ou inefficacement ces IRIF. Par conséquent, le signal γ -H2AX en utilisant un anticorps anti- γ H2AX ainsi que le nombre de *foci* de RAD51 (anticorps anti-RAD51) serviront de marqueurs de la fonctionnalité de la recombinaison homologue pour faire la distinction entre les lignées cellulaires sauvages ou déficientes en HR et par suite déficientes en BRCA1 ou BRCA2.

Dans cette étude Vaclová *et al.*, ils ont évalué l'effet de 15 variants de *BRCA1* à partir de lignées cellulaires lymphoblastoïdes du sang périphérique issues de quinze femmes porteuses de variation *BRCA1* à l'état hétérozygote (*BRCA1 +/-*) sur la recombinaison homologue.

Ces tests présentent l'avantage de tester les variants dans le contexte génomique du patient. Ils sont cependant dépendants de la disponibilité d'une lignée cellulaire du patient.

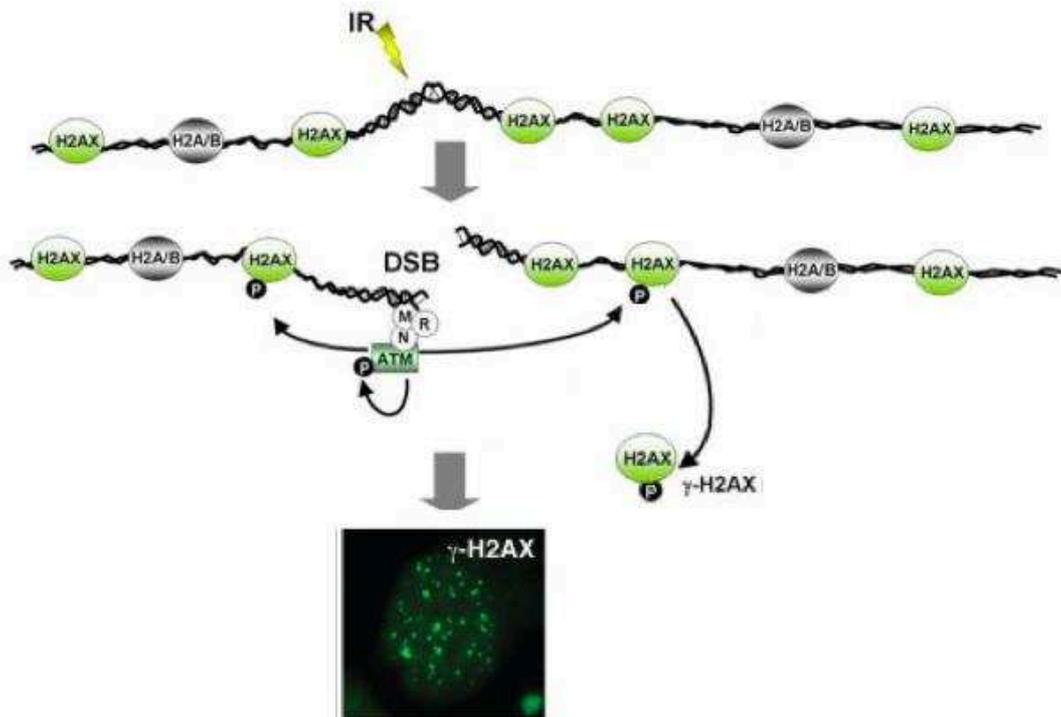


Figure 21 : Formation de *foci* γ -H2AX en réponse aux cassures double brin de l'ADN

Les radiations provoquent l'apparition des cassures double brin de l'ADN reconnues par le complexe MRN (MRE11-RAD50-NBS1), et qui vont favoriser le recrutement de la kinase ATM et son autophosphorylation sur le site lésé. ATM va phosphoryler l'histone H2AX. L'accumulation de cette protéine phosphorylée sous forme de γ -H2AX va produire des foyers détectables en immunofluorescence qui vont rapidement disparaître après réparation des lésions. Ces *foci* sont souvent utilisés comme traceurs de la présence de cassures double brin et comme marqueurs d'efficacité de la recombinaison homologe.

IR : Irradiation ; P : phosphate ; M : MRE11 ; N : NBS1 ; R : RAD50.

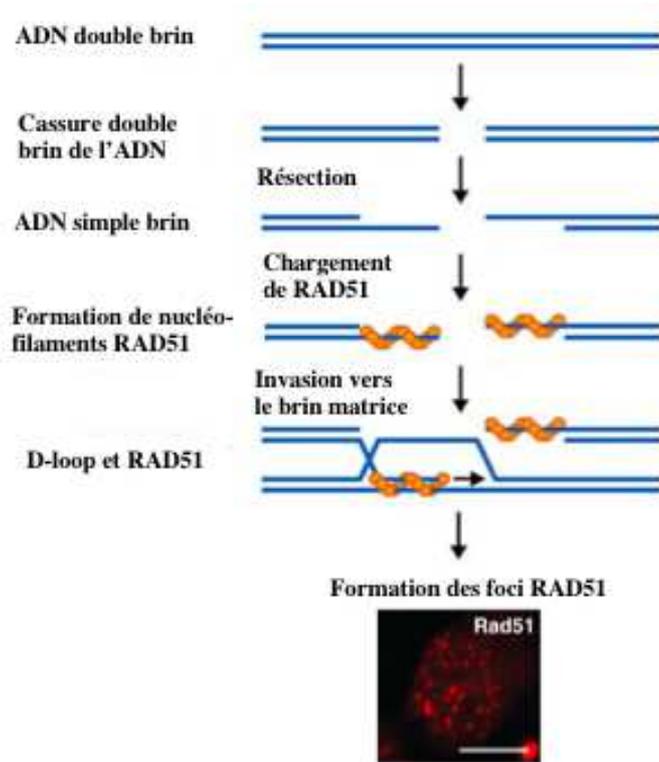


Figure 22 : Formation des *foci* RAD51 en réponse aux cassures double brin de l'ADN

Suite à un stress génotoxique entraînant des cassures double brin, RAD51 se lie aux zones d'ADN simple brin formées après la résection des extrémités lésées. Ensuite, il y aura assemblage de RAD51 sous forme de nucléofilaments, afin d'envahir la chromatide sœur et de former une D-loop. Cette accumulation de protéines RAD51 provoque la formation de foci détectables par immunofluorescence.

B- Tests basés sur l'amplification du centrosome

En effet, l'inactivation des gènes *BRCA1* ou *BRCA2* provoque une amplification du nombre de centrosomes. Cette amplification du nombre de centrosomes sera un marqueur de l'instabilité chromosomique et peut être utilisée comme mesure de la fonction des protéines BRCA, après la transfection des cellules par les formes sauvages ou mutantes marquées par *GFP* et soumises à une immunofluorescence directe en utilisant un anticorps contre la péricentrine, protéine clé du centrosome (Doxsey *et al.*, 1994). Ces essais ont été réalisés sur des cellules VC8 (*BRCA2-deficient VC8 cells*) (Wu *et al.*, 2005b). Une autre équipe a également utilisé ce test sur des cellules épithéliales mammaires non cancéreuses immortalisées (MCF-10A) en éditant leur génome afin de créer un panel de cellules isogéniques mutées pour *BRCA1* (Cochran *et al.*, 2015).

C- Tests domaines spécifiques

La protéine BRCA1 multi-domaines possède de nombreuses interactions moléculaires pour exercer des rôles distincts dans les processus cellulaires essentiels à la survie. Différents tests spécifiques ont été développés en se basant sur les fonctions portées par ces domaines.

Parmi ces tests, le test d'activation de la transcription (TA), qui évalue l'effet des variants touchant la région C-terminale de BRCA1, assurant une activité de transcription. Dans cet essai, une fusion de GAL4 ou LexA DBD, domaine de liaison à l'ADN des facteurs de transcription GAL4 et LexA, avec la région carboxy-terminale de BRCA1 (aa 1396-1863) contenant le variant à tester, est exprimée dans des cellules de levure ou de mammifère (Monteiro, 2000). Ces cellules ont des gènes rapporteurs intégrés dans leur génome ou sous une forme épisomale qui sont pilotés par des promoteurs contenant des sites de liaison GAL4 ou LexA. La quantification du produit du gène rapporteur ou de son activité permet une évaluation indirecte de l'activité transcriptionnelle induite par la protéine de fusion BRCA1 hétérologue. Le degré d'inactivation fonctionnelle associé à une altération de séquence donnée peut ensuite être évalué (Mirkovic *et al.*, 2004; Phelan *et al.*, 2005).

Un autre essai a été développé évaluant l'effet des variants touchant le domaine RING de BRCA1 qui est responsable de l'activité ubiquitine ligase et assure la médiation des interactions entre BRCA1 et BARD1. Dans ce test, ils ont évalué si les variants touchant ce

domaine spécifique pourront changer le niveau d'activité d'ubiquitine ligase (Stewart *et al.*, 2017).

D- Tests fonctionnels basés sur l'utilisation des cellules souches embryonnaires modifiées de souris

Cet essai repose sur la transfection du gène entier de BRCA1 ou BRCA2 humain sauvage ou muté, contenu dans des BAC (Bacterial Artificial Chromosome) dans des cellules qui contiennent un allèle défectueux pour l'un des gènes BRCA, et l'autre allèle fonctionnel mais pouvant être inactivé sous condition. L'impact de la variation à tester repose sur l'évaluation de la fonctionnalité de la protéine BRCA en observant la survie des cellules après inactivation de l'allèle sauvage restant. BRCA1 et BRCA2 sont essentielles au maintien et à la survie des cellules souches embryonnaires de souris (**Figure 23**) (Gowen *et al.*, 1996; Hakem *et al.*, 1996; Sharan *et al.*, 1997).

En effet, contrairement aux tests fonctionnels protéiques seuls qui utilisent des vecteurs d'expression portant uniquement la région codante (ADNc), cet essai repose sur la transfection du gène entier de BRCA1/2 humain sauvage ou muté, permettant ainsi d'appréhender l'impact de la variation sur l'épissage dans un contexte génomique proche du naturel. Ainsi, avec cette approche, il est possible d'étudier l'effet combiné des variations sur l'épissage et la protéine. De plus, tout type de mutation, y compris les mutations régulatrices et les altérations introniques, peut être introduit et évalué pour déterminer l'influence sur la survie des cellules souches embryonnaires (Kuznetsov *et al.*, 2008; Chang *et al.*, 2009; Bouwman *et al.*, 2013).

Cependant, il est important de noter que les systèmes basés sur l'expression hétérologue de BRCA humain chez la souris peuvent ne pas évaluer avec précision les variants situés dans des domaines présentant des niveaux de conservation inférieurs. De plus, ce système utilise une inactivation constitutive d'un allèle et une suppression conditionnelle de l'autre suite à l'introduction de BRCA1 ou BRCA2 humains. Notamment, en raison des multiples manipulations génétiques et des étapes de sélection, ces systèmes demandent beaucoup de travail et sont relativement lents. Par conséquent, ils risquent d'être très vite limités dans leur utilisation par le nombre faible de variants pouvant être analysés.

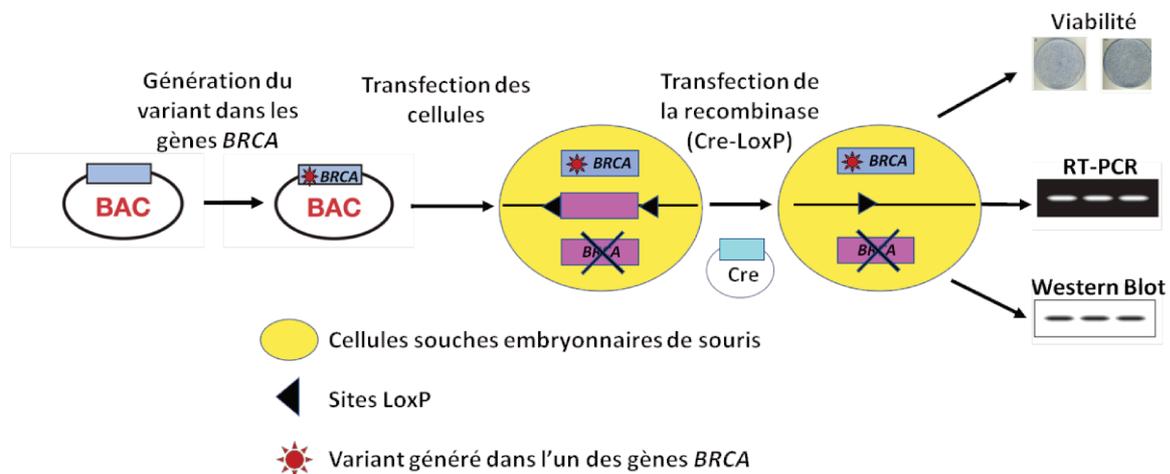


Figure 23 : Principe de l'essai fonctionnel basé sur l'utilisation des cellules souches embryonnaires modifiées de souris

Le gène entier de BRCA1 ou BRCA2 humain sauvage ou muté, contenu dans des BAC (Bacterial Artificial Chromosome) est transfecté dans des cellules souches embryonnaires de souris modifiées, contenant un allèle défectueux pour l'un des gènes BRCA et un autre allèle fonctionnel qui sera inactivé après transfection par la recombinaise Cre-LoxP. Ensuite, des analyses de viabilité cellulaire des cellules souches, de RT-PCR et de Western Blot sont effectuées afin d'évaluer la fonctionnalité des variants introduits.

E- Test à haut-débit

Des tests avec une plus grande capacité doivent donc être développés pour pouvoir répondre au très grand nombre de VSI de ces gènes qui va être généré avec la systématisation des panels de gènes. Récemment, une équipe américaine a développé un test fonctionnel à grande échelle en se basant sur la dépendance des cellules haploïdes d'une copie fonctionnelle du gène BRCA1 ou BRCA2 pour leur survie (Findlay et al., 2018). En effet, ils utilisent une modification du génome par le système CRISPR / Cas9 pour analyser tous les variants de nucléotides possibles (SNV) dans les exons codant des domaines fonctionnellement critiques de BRCA1. Leur essai repose sur la transfection de **cellules haploïdes humaines HAP123**, pour lesquelles la présence du gène BRCA1 est essentielle à la survie, par une construction co-exprimant Cas9 et des ARN guides (ARNg) ainsi qu'une librairie contenant les SNV à tester. Lors de la transfection, les cellules modifiées avec succès porteront un seul SNV de BRCA1. La viabilité cellulaire grâce à un test de survie basé sur la luminescence a été évaluée (**Figure 24**). Avec cette approche plus de 4000 SNV de 13 exons portant des domaines fonctionnels importants de BRCA1 ont été testés permettant l'identification de 821 variations délétères.

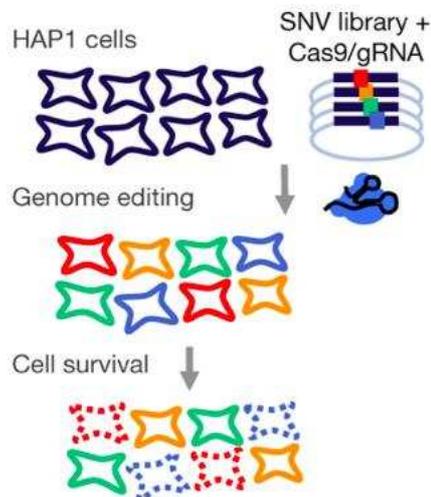


Figure 24 : Schéma illustrant les étapes du test à haut-débit (D'après Findlay et al., 2018)

Tous les SNV possibles ont été introduits dans le gène BRCA1 codant au niveau du domaine RING et BRCT. Pour chaque exon de BRCA1, une construction Cas9 / ARNg a été transfectée avec une banque de plasmides contenant tous les SNV. Cette librairie ainsi qu'une construction co-exprimant Cas9 et des ARN guides (ARNg) sont co-transfectés dans des cellules HAP1 haploïdes humaines, dont la survie est dépendante de la présence du gène BRCA1. L'effet des variants est évalué en analysant la survie des cellules.

L'édition du génome à haut-débit offre donc une alternative prometteuse pour tester à grande échelle tous les variants délétères dans ces gènes. Cette approche doit encore être appliquée pour caractériser un nombre appréciable de VSI dans BRCA1 ou dans d'autres gènes liés de manière similaire à la prédisposition au cancer.

Collectivement, ces études montrent la puissance des tests fonctionnels pour contribuer à la détermination de la pathogénicité des VSI de BRCA1 et BRCA2. Afin de surmonter les freins techniques rencontrés au moment de l'application de ces tests, de futures études pourraient s'appuyer sur l'utilisation de nouvelles techniques permettant d'analyser l'effet des mutations dans le contexte génétique des patients sur un prélèvement rapide non invasif. Une telle innovation devrait toutefois permettre de mener des analyses à plus haut-débit qui permettent, non seulement de classer efficacement les VSI mais également être informatives quant à la réponse thérapeutique.

À l'avenir, il sera important de considérer la meilleure façon d'incorporer les données provenant d'essais fonctionnels dans des modèles prédictifs multifactoriels qui incluent également toutes les données disponibles sur la co-ségrégation et la pathologie tumorale et sur les données génétiques et épidémiologiques. La puissance de prédiction de la pathogénicité des variants que ce type d'algorithme de calcul effectue peut-être très utile une fois que de grandes quantités de données fonctionnelles robustes sont disponibles.

II- Le syndrome de Li-Fraumeni (LFS)

Le syndrome de Li-Fraumeni (LFS) est l'une des formes héréditaires de cancer les plus remarquables de par l'étendue du spectre tumoral associé et la précocité de survenue des tumeurs. Cette pathologie se transmet d'une manière autosomique dominante (Strong et al., 1992) et résulte de mutations constitutionnelles à l'état hétérozygote du gène suppresseur de tumeurs TP53 (Tumor Protein p53) (MIM*191170) codant pour un facteur de transcription impliqué dans le maintien de l'intégrité du génome (Malkin et al., 1990; Srivastava et al., 1990).

I. Définition clinique du LFS

C'est en 1969 que Frederick Li et Joseph Fraumeni, publièrent une étude épidémiologique rétrospective dans laquelle ils avaient analysé 648 cas de rhabdomyosarcomes chez des enfants âgés de 0 à 14 ans (Li & Fraumeni, 1969). Ces enfants présentaient des tumeurs des tissus conjonctifs et avaient pour certains dans la famille des membres qui étaient atteints de cancer. Ils ont commencé ensuite à montrer qu'il y avait une certaine agrégation à partir d'une observation de 24 familles présentant une grande variété de cancers précoces (Li et al., 1988). Ces tumeurs associées à ce syndrome sont les sarcomes des tissus mous, les ostéosarcomes, les tumeurs cérébrales, les corticosurrénales, les carcinomes des plexus choroïdes, les cancers du sein pré-ménopausiques et les leucémies. Ainsi il y a eu une première définition du syndrome de Li-Fraumeni comme étant une des formes de prédisposition héréditaire aux cancers les plus sévères, associant un large spectre tumoral et à un âge précoce de développement des cancers avec un âge moyen d'apparition de la première tumeur estimé à 25 ans (Bougeard et al., 2015).

Le risque de développer des tumeurs primitives multiples est excessivement élevé chez les patients LFS. En effet 49 % des porteurs de mutation constitutionnelle de *TP53* développeront une deuxième tumeur après une période de 10 ans en moyenne, et chez certains patients, jusqu'à 6 cancers primitifs distincts ont pu être observés (Mai *et al.*, 2016). Chez les porteurs de mutations constitutionnelles du gène *TP53* les traitements du premier cancer peuvent favoriser la survenue d'un deuxième cancer lié au traitement du cancer précédant. Cela a été montré par plusieurs équipes décrivant par exemple une prévalence accrue de cancers solides dans le champ d'irradiation chez les patients LFS ayant reçu une radiothérapie par rapport à ceux qui ne l'ont pas reçue (Limacher *et al.*, 2001; Pierce & Haffty, 2011).

Conformément à ces constatations une étude a été menée récemment dans notre laboratoire dans le but de montrer la contribution des traitements anticancéreux génotoxiques au développement des tumeurs radio- ou chimio-induites et identifier les traitements les plus adaptés à ces patients LFS (Kasper *et al.*, 2018).

De par l'étendue du spectre tumoral et l'incidence élevée de tumeurs primitives multiples, la prise en charge médicale des patients LFS représente un défi. Depuis la description initiale du syndrome LFS (Li & Fraumeni, 1969) la définition du LFS est passée d'une forme pédiatrique délimitée par des critères stricts à une forme définie par des critères plus larges incluant un large spectre de tumeurs et des âges de survenue plus tardifs. Cette redéfinition a été notamment possible grâce aux efforts du groupe de travail français sur le LFS qui a défini et optimisé les critères nosologiques de ce syndrome appelés « critères de Chompret » (Chompret *et al.*, 2001; Bougeard *et al.*, 2008, 2015; Tinat *et al.*, 2009).

Notre équipe impliquée depuis 20 ans dans l'étude du syndrome de Li-Fraumeni (LFS), travaillant en collaboration avec le laboratoire de Génétique Moléculaire du CHU de Rouen qui réalise le diagnostic moléculaire du LFS au niveau national, a contribué à la relecture du LFS. Grâce à une étude rétrospective d'une cohorte composée de plus de 200 familles atteintes du syndrome de Li-Fraumeni, porteuses d'une altération hétérozygote du gène *TP53*, notre équipe a ainsi participé à redéfinir les critères nosologiques du LFS correspondant à la version 2015 des critères de Chompret, faisant référence aujourd'hui au niveau international (Bougeard *et al.*, 2015). Ces critères cités dans le tableau 2 illustrent la diversité d'expression clinique du LFS avec une présentation familiale, des tumeurs primitives multiples chez le patient, des tumeurs pédiatriques rares mais surtout un cancer du sein précoce chez la femme jeune.

CRITERES DE CHOMPRET Version 2015	Critères familiaux	Le cas index atteint d'une tumeur appartenant au spectre tumoral du LFS (sarcome des tissus mous, ostéosarcome, tumeur cérébrale, cancer du sein pré-ménopausique, corticosurréalome, leucémie ou cancer du poumon broncho-alvéolaire) avant l'âge de 46 ans ET au moins un apparenté du premier ou du deuxième degré atteint d'une tumeur du spectre étroit du LFS (sauf cancer du sein si le proposant est atteint lui-même d'un cancer du sein) avant l'âge de 56 ans ou présentant des tumeurs multiples
	Critères individuels	Le cas index présentant des tumeurs primitives multiples (sauf tumeurs mammaires multiples), dont deux appartenant au spectre tumoral du LFS et dont la première est survenue avant l'âge de 46 ans
	Tumeurs rares	Le cas index atteint d'un corticosurréalome ou d'une tumeur des plexus choroïdes ou un rhabdomyosarcome embryonnaire anaplasique , quels que soient l'âge et les antécédents familiaux
	Cancer du sein	Le cas index ayant développé un cancer du sein d'apparition précoce avant l'âge de 31 ans et chez qui la recherche de mutations des gènes <i>BRCA1/BRCA2</i> a abouti à un résultat négatif

Tableau 3 : Présentation des quatre portes d'entrée cliniques du syndrome de Li-Fraumeni d'après les critères de Chompret actualisés en 2015 (Bougeard *et al.*, 2015)

LFS = *Li-Fraumeni Syndrome* ou Syndrome de Li-Fraumeni

II. Bases moléculaires du LFS

L'étiologie moléculaire du LFS a été découverte en 1990 par les deux équipes de Stephen Friend (Malkin et al., 1990) et d'Esther Chang (Srivastava et al., 1990) et leurs collaborateurs par approche de gène candidat en se basant sur des observations antérieures d'inactivation somatique de TP53 dans les formes sporadiques de cancers tels que les tumeurs cérébrales, les cancers du sein, du poumon ou encore du colon (Nigro et al., 1989). Grâce à leur approche, ils ont démontré que les individus LFS hébergent des mutations constitutionnelles dans TP53, gène codant pour le facteur de transcription p53 (protéine tumorale p53). Depuis lors, près de 1229 mutations constitutionnelles ont été répertoriées à ce jour dans la base de données de l'IARC (International Agency for Research on Cancer, TP53 Database, R17, août 2018). A ce jour, aucun autre gène n'a pu être mis en évidence dans les familles évocatrices du LFS, faisant de *TP53* le seul gène impliqué dans le LFS.

a- Le gène TP53

Le gène TP53 est situé sur le chromosome 17p13.1 (NC_000017.10, GRCh37.p13) et s'étend sur environ 20 kb d'ADN génomique. Il est composé de 11 exons (NM_000546) dont dix qui codent pour la protéine p53 de 393 acides aminés, un exon non codant (exon 1) et trois exons alternatifs (exons 2/3, 9b, 9c) impliqués dans la synthèse de transcrits alternativement épissés codant pour les isoformes de la protéine p53.

b- La protéine p53

La protéine canonique p53 a été découverte en 1979 comme protéine cellulaire de 53 kDa co-immunoprécipitant avec le grand antigène T du virus SV40 dans des cellules transformées (Lane & Crawford, 1979; Linzer & Levine, 1979). Historiquement, il a été d'abord proposé que TP53 fonctionnait comme un oncogène, étant donné que de nombreuses tumeurs issues de sarcomes ou de leucémies, exprimaient des niveaux accrus de la protéine p53 (Rotter et al., 1983; Eliyahu et al., 1984). Cependant, des investigations plus poussées dans les années 1989 ont montré que les études concernaient la forme mutante et que p53 sauvage, codée par le gène TP53, supprime la croissance et la transformation oncogénique en culture cellulaire (Finlay et al., 1989), permettant de reclasser la protéine en tant que principal suppresseur de tumeur et non pas d'oncogène (Baker et al., 1989). Des études ont fourni la première description d'une séquence consensus de liaison à p53, permettant éventuellement une recherche computationnelle à l'échelle du génome pour des sites de liaison de p53 putatifs (Fields & Jang, 1990). Grâce à ces études et à de nombreuses études ultérieures (el-Deiry et al., 1992; Shiao et al., 1992; Farmer et al., 1992), p53 a été définie comme facteur de transcription au début des années 1990 (Levine & Oren, 2009).

▪ **Domaines protéiques de p53**

Cette protéine humaine de 393 acides aminés est composée de cinq domaines fonctionnels (**Figure 25**) :

- **Un domaine d'activation transcriptionnelle N-terminal** impliqué dans le recrutement de cofacteurs transcriptionnels tel que les protéines co-activatrices p300 et CBP (*Creb Binding Protein*) et incluant deux domaines d'activation de la transcription (*TAD*), TAD1 et TAD2, qui couvrent respectivement les résidus d'acides aminés 1-40 et 40-60, et une région hautement conservée entre les résidus 13 et 26 jouant un rôle

dans l'interaction avec l'oncoprotéine MDM2 (murine double minute 2) qui est une ubiquitine ligase provoquant la dégradation de p53 par le protéasome.

- **Une région riche en résidus proline** contenant cinq motifs PXXP (entre les résidus 63-97) connus pour être impliqués dans l'interaction protéine-protéine au cours de la transduction du signal, grâce à leur liaison au domaine *SH3* (*Src homology 3*) (Yu *et al.*, 1994). Ce domaine est impliqué dans les mécanismes de régulation de la stabilité et des fonctions pro-apoptotiques de p53.
- **Un domaine de liaison à l'ADN central** ou DBD (résidus 98-292) hautement conservé, interagissant directement avec l'ADN en reconnaissant un motif ADN situé en amont du site d'initiation de ses gènes cibles. Cette séquence d'ADN consensus contient deux monomères de 10 paires de bases, séparés par 13 paires de bases dont la séquence consensus est RRRCWWGYYY-RRRCWWGYYY. La structure du domaine central est composée d'une part d'un motif structural boucle-feuillet-hélice qui comprend la boucle L1, les feuillets β S2 et S2', une partie du feuillet β S10 ainsi qu'une hélice en C-terminal. D'autre part un deuxième motif structural est formé par les boucles L2 et L3 qui sont stabilisées par un atome de zinc au niveau des cystéines 176, 238 et 242 et de l'histidine 179.
- **Le domaine de tétramérisation** (résidus 324-355) permet la formation de dimères et de tétramères de p53, essentielle à son activation. En effet, chaque monomère se compose d'un feuillet β et d'une hélice α reliés entre eux par une glycine au codon 334. Deux monomères forment un dimère qui est stabilisé par dimérisation antiparallèle des feuillets β et hélices α . Ensuite deux dimères forment ensemble un tétramère qui est stabilisé par des interactions hydrophobes (**Figure 26**).
- **Le domaine carboxyl-terminal** (résidus 355-393) riche en lysine est également reconnu comme domaine de régulation, qui régule l'activité et la stabilité de p53 grâce à de multiples modifications post-traductionnelles. Il permet grâce à sa flexibilité structurale d'interagir avec un grand nombre de protéines comme les sous-unités du facteur de transcription TFIIH ou XPD et XPB (*Xeroderma Pigmentosum Groupe B or D*).
- En plus de ses 5 domaines structuraux, p53 contient un signal de localisation nucléaire (*NLS* : 316 à 325) ainsi qu'un signal d'export nucléaire (*NES*, *Nuclear Export Signal*) situé dans le domaine de tétramérisation (acides aminés 340 à 351). Les deux permettent de réguler la localisation subcellulaire de p53. Des mutations dans ces

domaines pourraient donc impacter la localisation de la protéine (Shaulsky *et al.*, 1990).

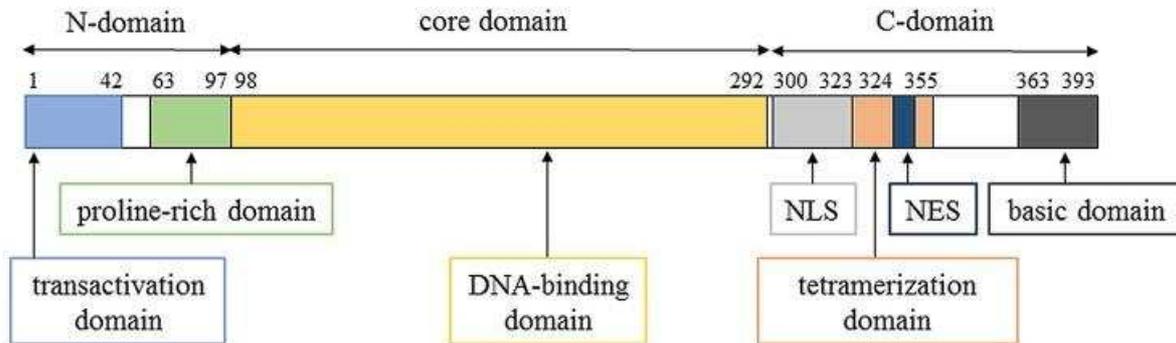


Figure 25 : Domaines de la protéine p53 humaine (D'après Tanaka *et al.*, 2018)

La protéine p53 humaine est une protéine multi-domaine composée de 393 résidus d'acides aminés et possède un domaine de transactivation, un domaine riche en proline, un domaine de liaison à l'ADN, un signal de localisation nucléaire, un domaine de tétramérisation, un signal d'exportation nucléaire et un domaine régulateur carboxyl-terminal.

NLS: Nuclear Localization Signal; NES: Nuclear Export Signal

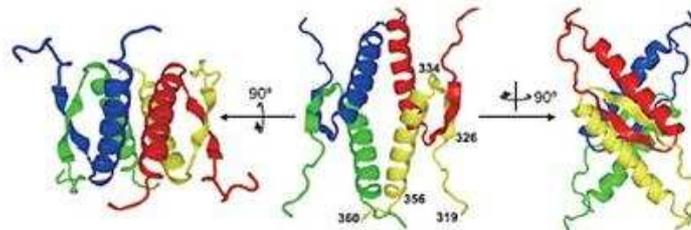


Figure 26 : Représentation de la structure cristallographique du domaine de tétramérisation de p53 dans 3 orientations différentes (D'après Tanaka *et al.*, 2018)

c- Fonctions cellulaires de p53

En tant que suppresseur de tumeur, la protéine p53 joue un rôle essentiel dans le maintien de l'intégrité du génome, dans la régulation de la prolifération et dans la régulation de l'apoptose (Kasthuber & Lowe, 2017).

La première fonction de la protéine p53 est d'agir en tant que **facteur de transcription**. Elle fonctionne en se fixant de manière spécifique sur les régions régulatrices de gènes dont elle contrôle l'expression. Les cellules aux cours des conditions physiologiques en absence de tout stress, contiennent des taux très faibles (souvent indétectables) de protéine p53 sauvage latente malgré une expression d'ARN messenger facilement détectable. La principale raison en est que p53 est la cible de régulateurs négatifs dont le principal est la protéine MDM2 possédant une activité E3 ubiquitine-ligase (Haupt *et al.*, 1997; Kubbutat *et al.*, 1997; Honda *et al.*, 1997) qui joue un rôle de régulateur négatif de p53 et dont l'expression est induite par p53. MDM2 est capable d'ubiquitiner l'extrémité C-terminale de p53 au niveau de résidus lysine favorisant ainsi sa dégradation par le protéasome, assurant dans les cellules normales le maintien d'un faible taux de protéine p53, avec une demi-vie estimée de 5 à 30 minutes (Moll & Petrenko, 2003). En réponse à divers stimuli de stress, comme l'activation d'oncogènes, les lésions de l'ADN ou la privation en nutriments, les niveaux de protéine p53 augmentent considérablement. En effet, en réponse aux facteurs de stress susmentionnés, une série de modifications post-traductionnelles (phosphorylation, acétylation, méthylation, ubiquitination et interaction avec des cofacteurs) sur la protéine p53 entraînera l'inhibition de MDM2 par le suppresseur de tumeur ARF. Cela aboutira à l'activation de plusieurs voies de signalisation et à la dissociation du complexe p53-MDM2. p53 est ainsi stabilisée, activée et ensuite transférée vers le noyau. Une fois dans le noyau, la région N-terminale libérée de p53 va s'associer aux co-activateurs p300/CBP, renforçant la stabilisation de p53, et provoquant un changement de conformation permettant à p53 de former des tétramères fonctionnels et de se lier par son domaine de liaison à l'ADN hautement conservé (DBD) à des séquences spécifiques d'ADN appelées élément de réponse à p53 (*p53RE*, *p53 Response Element*) (**Figure 27**).

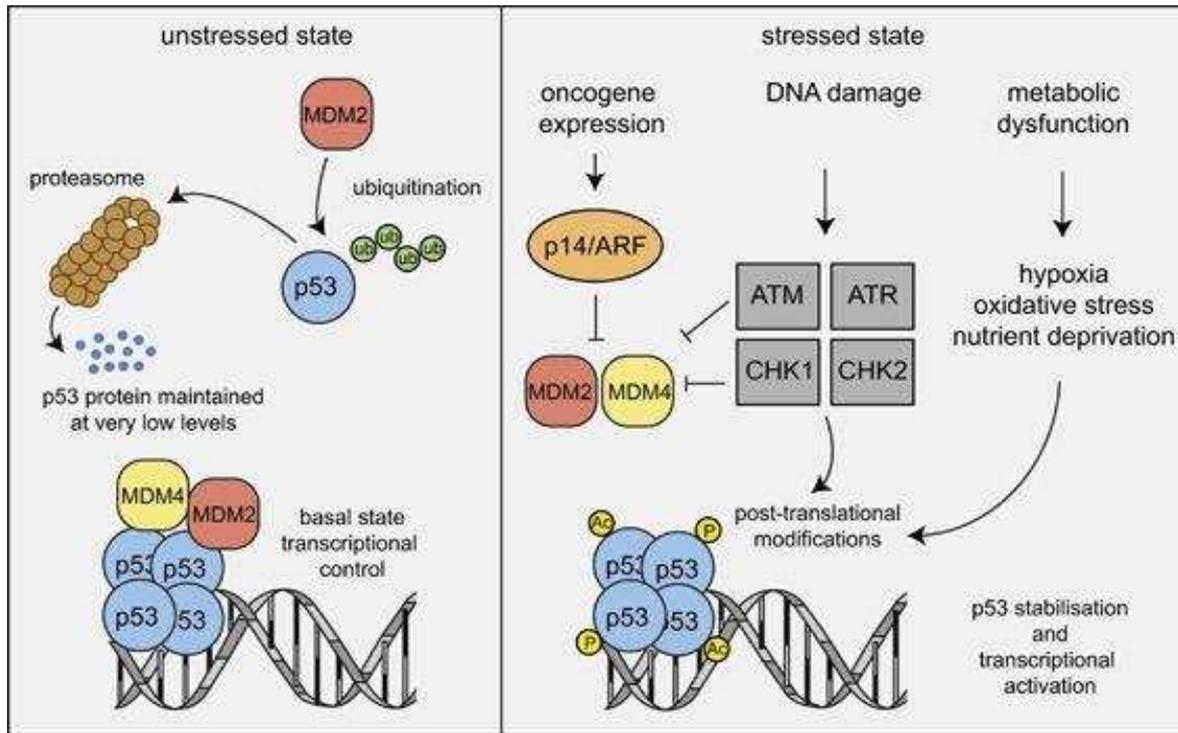


Figure 27 : Régulation du niveau et de l'activité de la protéine p53 dans les cellules non stressées par rapport aux cellules stressées (D'après Aubrey *et al.*, 2018)

Dans les cellules non stressées, la protéine p53 est maintenue à de faibles niveaux par une série de régulateurs comprenant MDM2 qui fonctionne comme une ubiquitine ligase pour faciliter la dégradation de p53 par le protéasome. Cependant, dans les cellules soumises à un stress (activation d'oncogènes, lésions de l'ADN ou dysfonctionnement métabolique), MDM2 sera inhibée, p53 va subir des modifications post-traductionnelles (acétylation, phosphorylation) bloquant ainsi sa séquestration par MDM2 et par la suite sa dégradation. p53 sera stabilisée et va se fixer sur les régions régulatrices de ses gènes cibles afin de jouer son rôle de facteur de transcription.

Ub, Ubiquitine ; P, phosphorylation ; Ac, Acétylation

Ces éléments de réponse sont situés généralement dans les promoteurs des gènes à des distances variables du site d'initiation de la transcription (*TSS*, *Transcription Start Site*) mais également dans les séquences introniques ou même dans des exons (Beckerman & Prives, 2010) (**Figure 28**). La fixation de p53 à ces éléments de réponse va favoriser une étape de remodelage de la chromatine afin d'ouvrir la chromatine et recruter le complexe de pré-initiation de transcription (PIC, PreInitiation Complex) permettant à p53 d'agir comme un facteur de transcription régulant directement l'activité d'un ensemble de gènes (~500) impliqués dans un large éventail de processus cellulaires (Beckerman & Prives, 2010; Fischer, 2017; Aubrey *et al.*, 2018; Nguyen *et al.*, 2018).

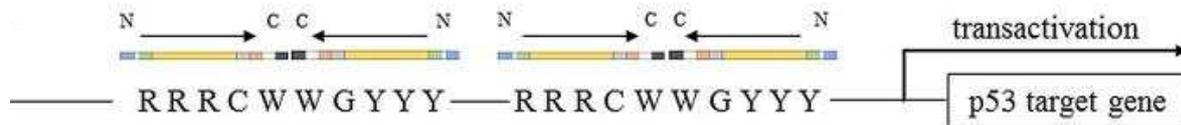


Figure 28 : Séquence consensus d'ADN des éléments de réponse à p53 (D'après Tanaka *et al.*, 2018)

La protéine p53 reconnaît la séquence RRRCWWGYYY (R : base purine, Y : base pyrimidine, W : adénine ou thymine) dans le noyau et se combine avec la séquence pour former un tétramère permettant ainsi l'activation de la transcription du gène cible.

Le facteur de transcription p53 est aussi capable d'activer et de réprimer d'autres gènes par **des interactions indirectes** :

- p53 peut inhiber l'expression de certains gènes via l'**activation d'une protéine répressive**. Par exemple, après des doses élevées d'irradiation γ , p53 transactive le gène *Slug*. La protéine *Slug* va se lier et réprimer le promoteur du gène *PUMA*, empêchant ainsi l'apoptose (Wu *et al.*, 2005a).
- Elle peut se lier à son élément de réponse et **bloquer la liaison d'un autre facteur de transcription** supprimant ainsi son activité. Par exemple, en réponse à un stress hypoxique, p53 se lie au promoteur d'AFP (Alpha Fetoprotéine) et déplace l'activateur transcriptionnel HNF3 (Hepatocyte Nuclear Factor 3) pour réprimer l'expression d'AFP (Lee *et al.*, 1999).

La protéine p53 peut aussi réprimer les gènes qui ne contiennent pas d'élément de réponse p53 via des **interactions protéine-protéine**. Le promoteur de *cycline B2*, par exemple,

contient des sites de reconnaissance du facteur de transcription NF-Y (Nuclear transcription Factor Y) mais pas de p53. NF-Y liée au promoteur interagit avec p53 et dans ce cas, p53 recrute HDAC1 (*Histone Deacetylase 1*) pour réprimer le promoteur de la cycline B2 (Imbriano *et al.*, 2005).

En fonction du stress cellulaire, de la nature de la cellule et de son environnement, l'activation de p53 va conditionner le destin de la cellule. En réponse à différents types de stress la protéine p53 va induire **l'arrêt du cycle cellulaire** aux points de contrôle de la fidélité de la division cellulaire pour éviter la propagation des lésions de l'ADN. En effet, p53 va activer la transcription de son gène cible *CDKN1A* (*Cyclin- Dependent Kinase Inhibitor 1A*) codant pour la protéine p21(Löhr *et al.*, 2003), inhibitrice des cyclines/cdk qui ne seront plus capables de phosphoryler la protéine Rb (*Retinoblastoma protein*). Rb hypophosphorylée piège le facteur de transcription E2F empêchant la transition **G1/S** (Xiong *et al.*, 1993; el-Deiry *et al.*, 1994; Niculescu *et al.*, 1998). p53 activée par des lésions de l'ADN est aussi capable de réprimer la transcription de Cdc25C (Cell Division Cycle 25C) (St Clair *et al.*, 2004) et transactive SFN (Stratifin) et GADD45A qui inhibent la formation du complexe cycline B1/Cdk1 en se liant à CDC2 (Cell Division Cycle Protein 2) (Hermeking *et al.*, 1997; Zhan *et al.*, 1999), résultant dans l'arrêt de la transition **G2/M** et permettant la suspension du cycle avant la mitose.

Cet arrêt peut être **un arrêt transitoire** permettant la **réparation des lésions**, processus dans lequel p53 joue également un rôle important. En effet p53 active différents gènes impliqués dans divers types de réparation de l'ADN, via son activité transcriptionnelle ou par des interactions protéine-protéine. p53 intervient dans la réparation par excision de nucléotide (NER) en transactivant les gènes *DDB2* (*DNA Damage-Binding protein 2*), *GADD45a* (*Growth Arrest DNA Damage 45*), *XPC* (*Xeroderma Pigmentosum group C*), *PCNA* (*Proliferating Cell Nuclear Antigen*) et le complexe RP-A (*Replication Protein A*) (Chen *et al.*, 1995; Abramova *et al.*, 1997; Adimoolam & Ford, 2002). Elle intervient également dans la réparation par excision de bases (BER) en réprimant l'expression de *MPG* (*N-Methylpurine DNA Glycosylase*), et en activant celle d'*APE1* (*Apurinic/aprimidinic endonuclease 1*) et d'*OGG1*(Zurer *et al.*, 2004), la réparation des mésappariements de l'ADN (MMR) en activant la transcription du gène *MSH2* (*MutS Homolog 2*) (Scherer *et al.*, 2000). Lors d'un stress réplicatif intense, p53 régule négativement la voie de recombinaison homologue (HR) grâce à des interactions directes avec les protéines RAD51, RAD54 et RPA, permettant de prévenir

un excès de recombinaison homologue qui serait alors néfaste pour le maintien de l'intégrité génomique (Arias-Lopez *et al.*, 2006 ; Menon & Povirk, 2014).

Lorsque les lésions sont trop importantes l'arrêt du cycle en phase G1 et G2 est **irréversible** conduisant :

- À la **sénescence prématurée** qui résulte de l'activation des protéines (PML Promyelocytic Leukemia Protein) (de Stanchina *et al.*, 2004), PAI-1 (Plasminogen Activator Inhibitor) (Andreasen *et al.*, 2000), du facteur de transcription DEC1 (*Deleted In Esophageal Cancer 1*) (Qian & Chen, 2008) et des gènes *BTG2* (B-Cell Translocation Gene 2) et *miR-34* (Tazawa *et al.*, 2007; Wheaton *et al.*, 2010), cibles de p53 impliqués dans la sénescence en réponse aux lésions de l'ADN et à l'exposition à des stress oncogéniques. Ce processus a pour but de prévenir le développement tumoral en stoppant de façon définitive la division des cellules endommagées à des stades précoces de transformation maligne malgré la présence de facteurs de croissance, limitant ainsi la durée de vie des cellules de mammifères et empêchant une prolifération illimitée.
- À la **mort cellulaire programmée** ou **apoptose**, à la fois intrinsèque et extrinsèque, pour éliminer les cellules trop endommagées en induisant la perméabilisation de la membrane externe mitochondriale par l'expression des protéines **pro-apoptotiques** (BAX : B-cell lymphoma 2 Associated X , BAK : B-cell lymphoma 2 Antagonist Killer, BBC3 : BCL2 Binding Component 3 , PMAIP1 (Phorbol-12-Myristate-13-Acetate-Induced Protein-1 ou NOXA), **anti apoptotiques** (BCL-2 B-Cell Lymphoma 2, BCL-xL, MCL1 Myeloid Cell Leukemia Sequence 1, BIRC3 Baculoviral IAP Repeat Containing 3 , BIRC5 ou Survivin) (Mihara *et al.*, 2003; Chipuk *et al.*, 2004; Leu *et al.*, 2004; Tomita *et al.*, 2006) et des **récepteurs membranaires à domaine de mort** (tels que CD95/Fas Fas Cell Surface Death Receptor), DR4 et DR5 (Death Receptor-4/5) et PERP (P53 Apoptosis Effector Related To PMP22) (Wu *et al.*, 1997; Bouvard *et al.*, 2000).
- À diverses formes de mort cellulaire non apoptotiques, y compris la **ferroptose** dépendant du fer et résultant des altérations du métabolisme ou de la disponibilité de ces acides aminés. p53 impacte la ferroptose à travers plusieurs mécanismes. En réprimant l'expression de SLC7A11, un composant clé de l'antiporteur cystine / glutamate, et la glutaminase 2, une des enzymes qui produisent le glutamate à partir de

la glutamine, la protéine p53 limite respectivement la capture de cystine et la production de glutamate, sensibilisant ainsi les cellules à la ferroptose induite par les dérivés réactifs de l'oxygène (*ROS, Reactive Oxygen Species*) (Jiang *et al.*, 2015; Labuschagne *et al.*, 2018).

- À l'induction de p53 l'**autophagie** anti-tumorale qui est un processus catabolique lysosomal de recyclage des constituants cellulaires, essentiel à l'homéostasie cellulaire. En réponse au stress, p53 va activer la transcription des gènes tels que *DRAM (Damage-Regulated Autophagy Modulator)* (Crighton *et al.*, 2006), *Ei24 (Autophagy Associated Transmembrane Protein)* (Zhao *et al.*, 2012) *SESN1/2 (Sestrin 1/2)* (Budanov, 2011). Plusieurs études ont identifié la réponse autophagique à l'activation de p53 comme étant importante pour la mort cellulaire médiée par p53. Par exemple, p53 favorise l'autophagie et la mort cellulaire des myocytes ventriculaires en activant la transcription de *Bnip3 (Bcl-2/adenovirus E1B 19-kDa Interacting Protein3)* (Wang *et al.*, 2013).

Lorsque l'expression de p53 n'est plus nécessaire, p53 stimulera l'expression de son propre régulateur négatif ce qui constitue une boucle d'autorégulation négative (Momand *et al.*, 2000). En maintenant l'intégrité du génome et en prévenant la prolifération des cellules avec un ADN endommagé, p53 agit comme "un gardien du génome" et joue un rôle clé contre la tumorigenèse.

Rôle de p53 dans le métabolisme

Au-delà des fonctions citées, p53 joue de multiples rôles dans le contrôle du métabolisme énergétique où elle exerce des activités qui contrecarrent le développement tumoral en s'opposant à l'effet Warbourg, plus précisément par sa capacité à diminuer la glycolyse utilisée préférentiellement comme source d'énergie par les cellules tumorales. (Labuschagne *et al.*, 2018). p53 est capable d'activer mais également de réprimer la transcription de nombreux gènes clés du métabolisme, modulant in fine **(i) la voie du glucose** (GLUT1, GLUT4, GLUT3 Transmembrane Glucose Transporter, RRAD Ras-related Associated with Diabetes, PGM PhosphoGlycerate Mutase) (Schwartzberg-Bar-Yoseph *et al.*, 2004; Kawauchi *et al.*, 2008; Zhang *et al.*, 2014), **(ii) des pentoses phosphates** (PFK PhosphoFructoKianse, TIGAR (TP53-Induced Glycolysis and Apoptosis Regulator) (Ros *et al.*, 2017), **(iii) le métabolisme des lipides** (p53 peut inhiber le PPP (Pentose Phosphate

Pathway) oxydatif, principale source de NADPH (Nicotinamide Adenine Dinucleotide Phosphate) pour la synthèse des acides gras, conduisant à une diminution de la lipogenèse (Jiang et al., 2011)) ou (iv) celui de certains **acides aminés non essentiels**, tels que la glutamine ou la sérine.

p53 active également de manière transcriptionnelle plusieurs autres cibles qui régulent négativement la voie mTOR (mechanistic Target Of Rapamycin), y compris AMPK (AMP-Activated Protein Kinase) via les sestrines 1 et 2 (SESN1/2), PTEN (Phosphatase and TENSin homolog) , et TSC1 / 2 (Tuberous Sclerosis Complex) (Budanov & Karin, 2008; Hasty et al., 2013). Ces réponses p53 limitent la croissance cellulaire et, de concert avec l'inhibition de la progression et de la prolifération du cycle cellulaire, permettent de préserver les ressources dans des conditions limitant les nutriments (Humpton & Vousden, 2016).

En parallèle, dans le but de s'opposer à l'effet Warburg, p53 stimule la **phosphorylation oxydative** ou la **respiration mitochondriale** par rapport à la glycolyse aérobie. En réponse à un stress oxydatif induit par ROS, p53 induit l'expression de *SCO2* (*Synthesis of Cytochrome c Oxidase 2*), *AIF* (*Apoptosis- Inducing Factor*), et *GLS2* (*mitochondrial Glutaminase 2*) (Suzuki et al., 2010; Itahana & Itahana, 2018) qui favorisent le cycle de Krebs.

D'autres formes de stress peuvent également activer p53, y compris une batterie de signaux regroupés sous le terme de « stress métabolique ». La capacité générale de p53 à aider les cellules à s'adapter et à survivre dans ces conditions d'épuisement en nutriments a récemment été décrite. L'activation des voies cataboliques telles que l'autophagie et l'oxydation des acides gras par p53 peut également aider à maintenir temporairement les cellules en l'absence de nutriments exogènes en favorisant la génération d'ATP. p53 semble jouer un rôle clé dans la réponse à ce manque sévère en nutriments en mobilisant des sources de nutriments et des voies cataboliques alternatives pour maintenir la survie des cellules tout en jouant un rôle dans l'élimination des cellules qui ne peuvent pas récupérer. Ces actions de p53, jouent un rôle important, à la fois pour les fonctions de suppresseur de tumeurs de cette protéine, mais également dans l'homéostasie tissulaire.

Autres fonctions de p53

En plus de son rôle dans la réponse biologique à différents stress, p53 régule divers aspects du comportement **des cellules souches**, notamment la **prolifération**, **l'auto-renouvellement**, et la **différenciation** (Olivos & Mayo, 2016). Dans ce contexte, il a été montré que p53 se lie au promoteur de Nanog, un gène nécessaire à l'auto-renouvellement dans les cellules souches embryonnaires de souris et réprime l'expression de Nanog après des lésions de l'ADN (Lin et al., 2005). Cela permettra à p53 de maintenir la stabilité génétique dans ces cellules, en induisant leur différenciation en d'autres types de cellules pouvant subir un arrêt du cycle cellulaire et une apoptose efficace. Une autre étude a montré que p53 limite la reprogrammation cellulaire, comme la génération de cellules souches pluripotentes induites (iPSC Induced Pluripotent Stem Cells) à partir de cellules différenciées, confortant ainsi l'idée qu'un aspect de la fonction de suppression de tumeur p53 consiste à restreindre la plasticité et à favoriser la différenciation (Krizhanovsky & Lowe, 2009). L'équipe de Stripp a aussi montré que p53 régule à la fois la prolifération et la différenciation des progéniteurs dans l'épithélium des voies respiratoires. Dans cette étude, la perte de p53 a diminué la différenciation des cellules ciliées et a amplifié le pouvoir de renouvellement automatique et la capacité de prolifération des cellules progénitrices, augmentant ainsi la densité des cellules épithéliales. Inversement, une copie supplémentaire de p53 augmenterait la quiescence et la différenciation des cellules ciliées (McConnell et al., 2016).

p53 est également impliquée dans la **migration cellulaire** suggérant un rôle potentiel de p53 dans le ralentissement de la progression maligne. Le mécanisme par lequel p53 inhibe la migration consiste à atténuer la transition épithélio-mésenchymateuse (EMT Epithelial-mesenchymal transition) en transactivant miR-200c et miR34a, des micro-ARN capables de neutraliser les conducteurs de la transition EMT, notamment Zeb1 et Snail (Muller et al., 2011; Chang et al., 2011; Kim et al., 2011). Des études supplémentaires sont nécessaires pour mieux comprendre la fonction de p53 dans ces processus.

p53 a la capacité d'inhiber l'**angiogenèse** par différents mécanismes, telle que la transactivation de la thrombospondine1 ou de Ptpn14 atténuant ainsi la croissance tumorale et les métastases (Bergers & Benjamin, 2003; Mello et al., 2017). Elle dirige aussi des changements dans le **microenvironnement tumoral**. Par exemple, p53 stimule les cellules immunitaires qui s'infiltrent dans les tumeurs et contribuent à leur élimination (Lujambio et al., 2013). p53 maintient aussi la **stabilité épigénétique** en contribuant à l'homéostasie de la

méthylation de l'ADN et préserve l'homogénéité clonale. Ce rôle a été montré grâce à l'observation des cellules souches embryonnaires de souris déficientes en p53 qui présentaient une méthylation de l'ADN globalement élevée ainsi qu'une hétérogénéité intra-clonale accrue (Tovy et al., 2017). La diversité des processus cellulaires régulés par p53 explique probablement sa puissante activité de suppresseur de tumeurs.

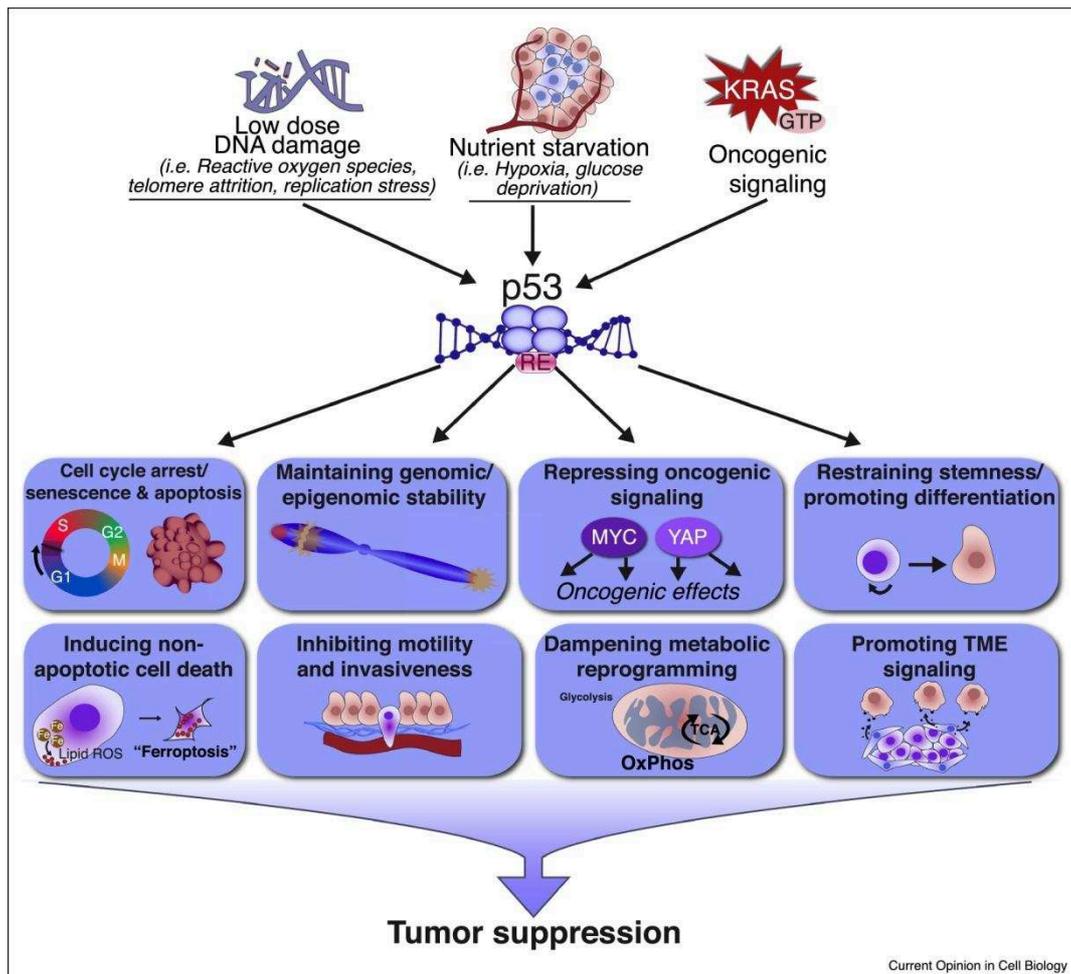


Figure 29 : Rôles de la protéine p53 (D'après Mello & Attardi, 2018)

En réponse aux différents types de stress cellulaire, p53 va être activée et va se lier à des éléments de réponse à l'ADN (RE) spécifiques afin de moduler différents processus cellulaires, impliqués dans la suppression des cellules tumorales.

d- Mutations constitutionnelles de TP53

i. Incidence et répartition des mutations

La base de données sur TP53 de l'International Agency for Research on Cancer (IARC) conserve un répertoire bien établi de mutations TP53 somatiques et constitutionnelles. D'après *l'IARC TP53 Database, R18, April 2016* (www-p53.iarc.fr), les mutations constitutionnelles du gène TP53 sont en très grande majorité (environ 73,6 %) des mutations faux-sens ce qui le diffère donc d'autres gènes suppresseurs de tumeurs tels BRCA1 ou BRCA2 qui sont majoritairement inactivés par des délétions ou des mutations non-sens aboutissant à des protéines tronquées (Hainaut & Hollstein, 2000; Petitjean et al., 2007) (**Figure 30**).

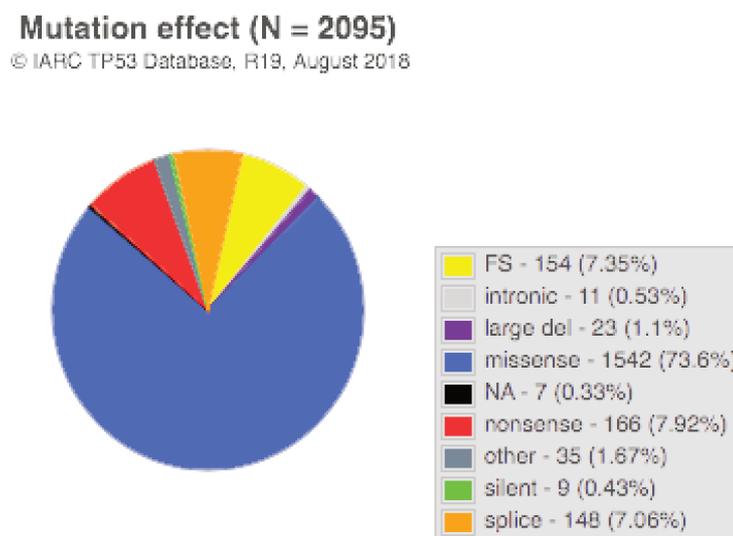


Figure 30 : Proportion des différents types de mutations constitutionnelles du gène TP53 (D'après IARC TP53 Database, R19, Août 2018)

Les mutations constitutionnelles les plus décrites sont les mutations faux-sens.

Les mutations sont réparties tout au long du gène, mais principalement localisées au niveau des exons 4 à 8 codant le domaine de liaison à l'ADN et plus particulièrement dans des régions très conservées au cours de l'évolution (**Figure 31**). Près de 15 % des mutations constitutionnelles de TP53 détectées chez les patients Li-Fraumeni sont des **mutations de**

novo et un cinquième de ces mutations de novo sont survenues pendant le développement embryonnaire et sont présentes sous forme de **mosaïques** (Renaux-Petel et al., 2018). De façon intéressante, les études menées en population générale ou sur des individus avec cancer non sélectionnés pour l'âge de survenue ou l'histoire familiale montrent que 0,2 % des individus portent une mutation prédite délétère de TP53 suggérant un gradient de sévérité des mutations TP53 (Andrade et al., 2017; Katki et al., 2018).

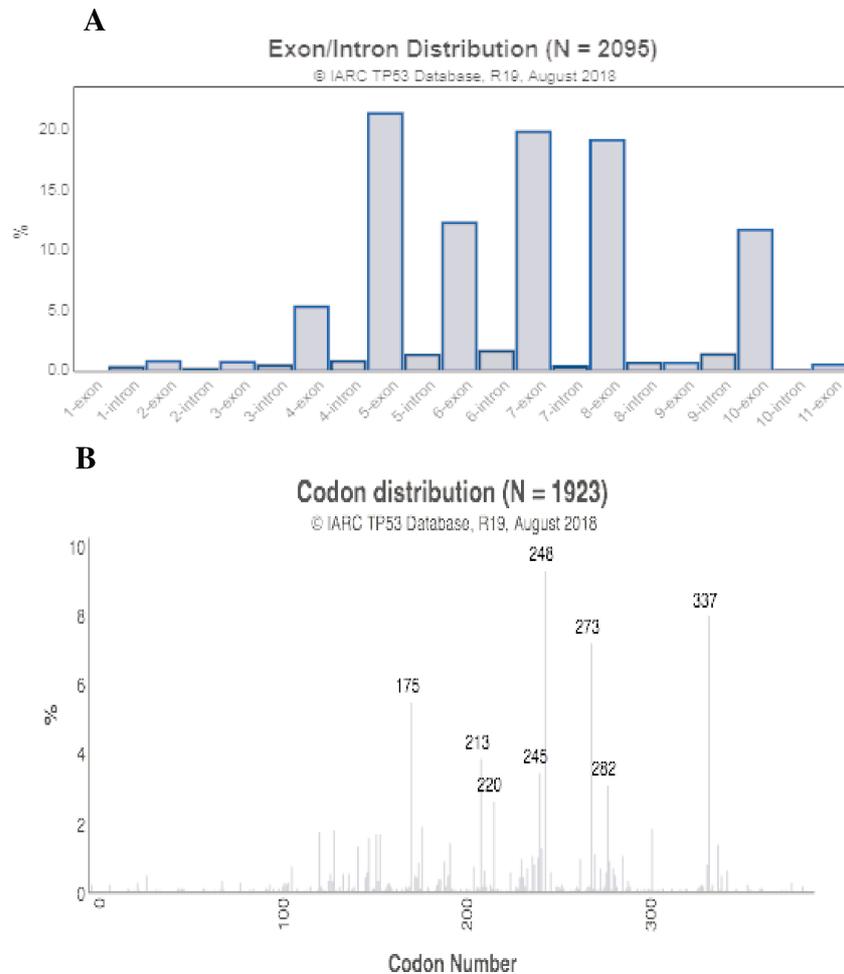


Figure 31 : Distribution des mutations constitutionnelles du gène *TP53* (D'après *IARC TP53 Database, R19, Août 2018*)

- A- Distribution des mutations de *TP53* le long du gène
- B- Les codons les plus fréquemment mutés au niveau de *TP53*

ii. Impact fonctionnel des mutations

Il existe plusieurs types d'effets par lesquels une mutation constitutionnelle de *TP53* peut favoriser la formation de tumeurs :

- **La perte de sa fonction** de suppresseur de tumeur (LoF) provoquée par une expression réduite de p53 ou par la perte totale de son activité transactivatrice et qui est l'effet commun des mutations du gène TP53 (Bougeard et al., 2003; Kato et al., 2003). Une expression réduite peut résulter de délétions génomiques qui éliminent tout ou une partie de TP53. Les mutations faux-sens de TP53 peuvent affecter ses domaines fonctionnels ou déstabiliser cette protéine globulaire. La majorité des mutations constitutionnelles de TP53 sont regroupées dans le domaine de liaison à l'ADN hautement conservé. Ces mutations produisent des protéines p53 mutantes avec une liaison à l'ADN altérée et donc une régulation altérée des gènes cibles en aval. Parmi les hotspots de mutations, on distingue (i) les mutants dits « de contact » impliquant les codons (arginines 248 et 273) qui correspondent aux sites de fixation de la protéine p53 à l'ADN ; une mutation de l'un de ces codons entraînera donc une moindre fixation de la protéine au niveau de sa séquence consensus et ainsi une baisse de l'activité transcriptionnelle et (ii) les mutants dits « de conformation » ou « de structure » impliquant les codons (arginines 175, 249 et 282 et Glycine 245), qui n'entrent pas directement en contact avec l'ADN, mais qui vont permettre la stabilisation de la protéine p53 (Bullock & Fersht, 2001).

- Les mutations de type faux-sens dans *TP53* peuvent également exercer **un effet dominant négatif** dans lequel les monomères de la protéine p53 mutante interagissent physiquement avec ceux de la protéine p53 sauvage et donc inhibent l'activité de la protéine sauvage issue de l'allèle restant. Les chercheurs ont démontré que la présence du mutant p53 réduisait considérablement la capacité de la protéine de type sauvage à se lier aux promoteurs de gènes cibles sensibles à p53 (Hinds *et al.*, 1990; Milner *et al.*, 1991; Aramayo *et al.*, 2011). En effet, l'effet dominant négatif pourrait expliquer pourquoi certains porteurs de mutation de *TP53* présentent une rétention de l'allèle *TP53* de type sauvage dans leurs tumeurs. À cette fin, Varley *et al.*, 1997 ont démontré que seulement 57 % des tumeurs provenant de patients présentant des mutations constitutionnelles de *TP53* présentaient une perte de l'allèle de type sauvage et une rétention de l'allèle muté. L'inactivation de la protéine sauvage par la protéine mutante semble tributaire de la stabilité protéique et du ratio protéine sauvage/protéine mutante au sein d'un tétramère. Ainsi, une étude a suggéré que, dans le cas où les 2 formes

sont exprimées de manière égale, 3 types différents de tétramères peuvent se former : 25 % contenant uniquement la forme sauvage de la protéine, 50 % mixtes et 25 % contenant uniquement la forme mutée. Cela aboutit à l'obtention de 75 % tétramères non fonctionnels, expliquant l'effet plus drastique et l'âge de développement des tumeurs plus précoce chez les patients porteurs de mutations à effet dominant négatif que celui des mutations perte de fonction (Aramayo *et al.*, 2011; Bougeard *et al.*, 2015).

- L'effet « **gain de fonction** » **GOF** est caractérisé par l'acquisition de nouvelles propriétés par le mutant *TP53*, indépendamment de la présence de la protéine p53 de type sauvage (Olive *et al.*, 2004). Les premières preuves soutenant la fonction oncogénique du mutant *TP53* ont été dérivées de l'observation selon laquelle l'introduction du mutant *TP53* dans des cellules nulles *Trp53*^{-/-} permettait une croissance tumorale chez la souris. Les cellules complètement dépourvues de p53 forment des tumeurs qui régressent spontanément. En revanche, les cellules exprimant le mutant *TP53* présentent une excroissance mortelle (Shaulsky *et al.*, 1991; Dittmer *et al.*, 1993).

La protéine p53 mutée médie l'effet GOF par plusieurs mécanismes dont le plus important est sa capacité à se lier et d'inactiver les protéines p63 et p73 empêchant leur capacité à amortir la prolifération cellulaire et à promouvoir l'apoptose (Lang *et al.*, 2004). Un autre mécanisme important implique la modulation de gènes cibles tel que la régulation positive d'ABCB1 (ATP- Binding Cassette Subfamily B Member 1) d'EGR1 (Early Growth Response 1) et de NFkB2 (Nuclear Factor Kappa B Subunit 2) empêchant l'apoptose et contribuant à la résistance aux médicaments, de MYC (ou c-myc) activant la prolifération, et de TERT (Telomerase Reverse Transcriptase) favorisant la réplication et des gènes cibles de NFkB impliqués dans l'inflammation et l'angiogenèse. De plus, p53 mutée pourra réguler négativement des gènes comme ID2 (Inhibitor of DNA binding 2) et ATF3 (Activating Transcription Factor 3) (Brosh & Rotter, 2009). p53 mutée interagit également avec des protéines anti-tumorales (par exemple MRE11) en changeant leur fonction (Liu *et al.*, 2010), ou avec des facteurs de transcription (comme ETS2) favorisant le développement tumoral (Zhu *et al.*, 2015) (**Figure 32**).

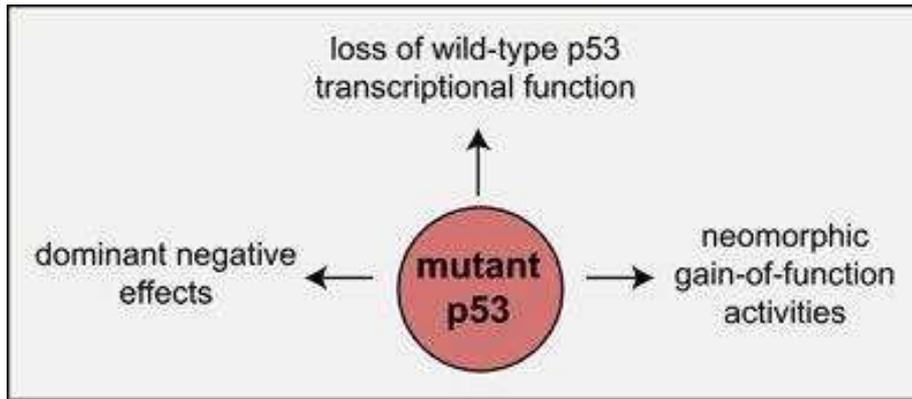


Figure 32 : Impact fonctionnel des mutations du gène *TP53* (D'après Aubrey et al., 2018)

Les mutations du gène *TP53* peuvent entraîner une perte de fonction partielle de p53. D'autres types de mutations, produisent des protéines mutantes capables d'inhiber, à des degrés divers, la fonction de la protéine de type sauvage codée par le second allèle. Les autres mutations confèrent de nouvelles propriétés de gain de fonction à p53 indépendantes de p53 de type sauvage.

III. Test fonctionnels dédiés à l'analyse des variations de *TP53*

Malgré l'accumulation de connaissances sur la fonction du gène *TP53*, la classification fonctionnelle des variants détectés dans ce gène reste un réel défi. De plus, il reste aujourd'hui impossible de prédire avec certitude quand et quels porteurs de mutation de *TP53* développeront un premier cancer ou un cancer ultérieur, comment retarder ou prévenir ces cancers et comment les traiter le plus efficacement possible. Plusieurs études génotype-phénotype impliquant des familles de LFS ont suggéré que le type de mutation constitutionnelle de *TP53* pourrait prédire le type de tumeur et l'âge du cancer (Birch et al., 1998). L'étude effectuée au sein de notre laboratoire (Bougeard et al., 2015) a montré un âge au diagnostic de la première tumeur significativement plus précoce chez les patients présentant des mutations faux-sens (âge moyen du début de la tumeur 23 ans) que chez ceux présentant des délétions/insertions, réarrangements génomiques (28 ans). La différence était encore plus grande lorsque les mutations faux-sens étaient comparées à des réarrangements génomiques correspondant à des formes extrêmes d'altérations nulles (35 ans). De plus, les patients dont les mutations étaient connues pour exercer un effet DN présentaient un âge encore plus précoce au début du cancer (21 ans).

Pour permettre de classer les variants de *TP53* et mieux étudier la corrélation génotype-phénotype, plusieurs essais fonctionnels ont été développés. La disponibilité de tests fonctionnels fiables pourrait aider à expliquer les diverses manifestations du LFS comme l'hétérogénéité phénotypique, la pénétrance variable, et les disparités dans l'âge d'apparition de la première tumeur et pourrait permettre de mettre en place des protocoles de surveillance individualisés.

Afin d'évaluer l'effet des différentes mutations constitutionnelles de *TP53* et la capacité de réponse des protéines mutantes au stress cellulaire et à induire l'expression de ses gènes cibles, plusieurs tests ont été développés. Des tests fonctionnels, relativement anciens, basés sur l'utilisation des lymphocytes ou fibroblastes en culture ont été mis en place. Ces essais se basaient sur l'analyse de la réponse de p53 suite à un stress génotoxique (lésions de l'ADN). Certains analysaient la capacité de la protéine mutante à provoquer l'arrêt du cycle cellulaire en phase G2 (Goi et al., 1997), d'autres analysaient l'instabilité chromosomique (Boyle et al., 1998), ou la capacité de la protéine mutante à induire l'apoptose (Camplejohn et al., 1995; Lomax et al., 1997), ou encore l'activité de transactivation (Fields & Jang, 1990; Raycroft et al., 1990; Farmer et al., 1992; Frebourg et al., 1992). Ces derniers sont basés sur l'utilisation de

systèmes indicateurs utilisant un gène rapporteur tel que la luciférase ou la beta-galactosidase avec un promoteur défectif en amont duquel sont clonées des séquences cibles de p53. L'ADN complémentaire de p53 sauvage ou mutant est alors exprimé dans ces systèmes et la capacité de la protéine mutante à transactiver le gène rapporteur est alors comparée à celle de la protéine sauvage. Ces tests fonctionnels de p53 humain, qui nécessitent un travail laborieux (incluant des étapes de clonage, de mutagenèse dirigée et de transfection), ont été développés dans la levure, présentant un excellent modèle pour l'étude des fonctions de la protéine p53 humaine, compte-tenu de l'absence du gène TP53 chez la levure et de la puissance de la recombinaison homologue dans la levure facilitant le clonage des ADNc dans un vecteur d'expression (Schärer & Iggo, 1992; Ishioka et al., 1993; Flaman et al., 1995; Lomax et al., 1997).

1- Analyse fonctionnelle des allèles séparés dans la levure ou FASAY (Functional Analysis of Separated Alleles in Yeast)

Brièvement, une souche mutante de levure *Saccharomyces cerevisiae* est co-transformée par les ADNc de p53 amplifiés par PCR à partir des ARN extraits des lymphocytes de patients LFS ou d'une simple biopsie tumorale, et par un vecteur d'expression. Cette souche inductrice est mutante pour le gène ADE2 impliqué dans la voie de biosynthèse des purines. La mutation de ce gène entraîne un blocage métabolique et l'accumulation d'un intermédiaire en amont le P-ribosyl-amino-imidazole (AIR) qui présente une pigmentation naturelle conférant à la levure la couleur rouge visible sur des boîtes de pétri sélectives. La souche est donc spontanément rouge lorsqu'elle est cultivée sur un milieu pauvre en adénine. Dans cette souche de levure ADE2-, un plasmide indicateur de l'activité transcriptionnelle de TP53 a été intégré. Ce plasmide indicateur contient le gène ADE2 sauvage sous le contrôle transcriptionnel d'un promoteur défectif CYC1 en amont duquel 3 sites de fixation de p53 à l'ADN correspondant à la séquence RGC (Ribosome Gene Cluster) ont été clonés. Cette séquence fut la première séquence de fixation de p53 identifiée qui n'a pas de rôle dans les voies biologiques contrôlées par TP53 et qui est utilisée par toutes les équipes pour étudier la fonction transcriptionnelle de TP53. Après recombinaison homologue (Gap Repair), l'expression de la protéine p53 sauvage dans cette souche indicatrice entraîne une levée du bloc métabolique et la levure devient blanche. En revanche l'expression de la protéine p53 mutante incapable d'induire l'expression du gène ADE2 génère des colonies rouges (**Figure 33**).

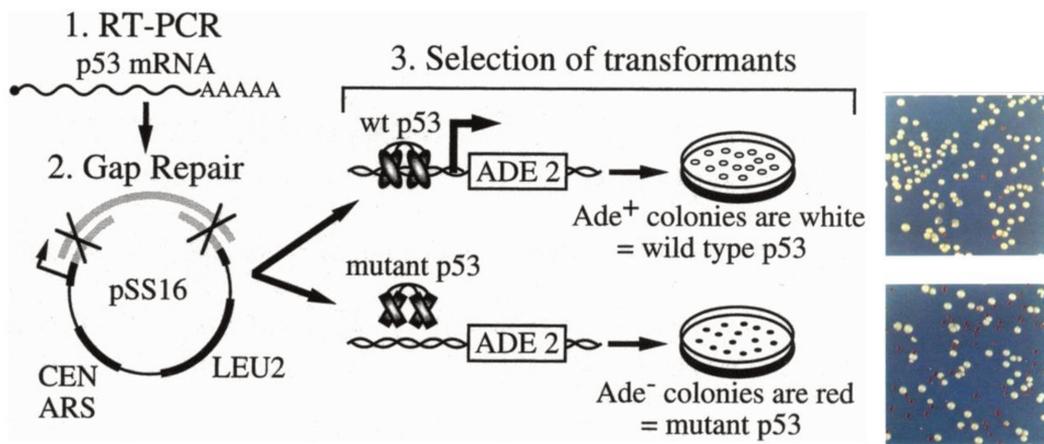


Figure 33 : Principe de la technique FASAY (D'après Flaman et al., 1995)

Les ARN totaux sont extraits des cellules, rétrotranscrits puis les ADNc de p53 sont amplifiés par PCR. Une souche mutante de levure *Saccharomyces cerevisiae* est alors co-transformée par les ADNc p53 et par un vecteur d'expression, dans lequel a été intégré le gène codant pour la leucine. Après recombinaison homologue, les levures ayant intégré le plasmide sont sélectionnées sur un milieu dépourvu de leucine. La souche de levure présente également une mutation du gène *Ade2* codant pour une enzyme impliquée dans le métabolisme de l'adénine. L'inactivation de cette enzyme induit l'accumulation d'un métabolite intermédiaire de couleur rouge. Ainsi, si la protéine p53 intégrée est sauvage, le métabolisme de l'adénine est restauré et les colonies sont blanches. A l'inverse, si la protéine p53 n'est pas fonctionnelle, elle ne peut pas activer la transcription du gène *Ade2* et les colonies sont rouges.

Cet essai fonctionnel a été utilisé en routine dans de nombreux laboratoires en tant qu'outil de diagnostic pour identifier des mutations somatiques ou constitutionnelles de *TP53*. Cette souche indicatrice a également représenté un outil puissant pour identifier des molécules capables de restaurer l'activité transcriptionnelle de p53 mutants. Cet essai simple, rapide, efficace et sensible permet de détecter l'effet des variants constitutionnels de *TP53* notamment dans les familles LFS. Il a permis de démontrer les conséquences fonctionnelles des mutants de type faux-sens canoniques de *TP53* qui ont pour effet une perte de fonction transcriptionnelle. Dans ce contexte, ce test a permis de mettre en évidence l'existence d'un gradient dans l'effet des mutations sur la transactivation des gènes cibles de p53, comme par exemple la mutation p.(Arg175Pro) (NM_00546.5) du gène *TP53* qui conservait la capacité d'activer *p21* mais pas *BAX* (Flaman *et al.*, 1998). Il a également contribué à la détection de phénotypes intermédiaires en évaluant l'effet des mutations induisant une sensibilité à la température qui produisait des colonies roses en cultivant les levures à différentes températures (Flaman *et al.*, 1995).

2- Test fonctionnel à grande échelle des mutations faux-sens de TP53 dans la levure (Test de Kato)

Dans le but de documenter la capacité de transactivation de p53, Kato et ses collaborateurs ont exploré une version améliorée du test *FASAY* en construisant par une technique de mutagenèse dirigée associée à un test fonctionnel à base de levure, 2314 mutants p53 distincts, qui représentaient toutes les substitutions possibles d'acides aminés provoquées par une mutation ponctuelle dans la protéine p53 (Kato *et al.*, 2003). Ensuite, l'impact de ces protéines p53 mutantes a été évalué en quantifiant leur activité transcriptionnelle. Brièvement, les ADNc de *TP53* avec les différentes mutations faux-sens ainsi que la forme sauvage de *TP53* ont été générés par des PCR en deux étapes. Chacune des mutations a été intégrée par recombinaison homologue dans un vecteur contenant un site de liaison à p53 (p21WAF1) sous la dépendance du gène rapporteur EGFP (*Enhanced Green Fluorescent Protein*) ensuite l'ensemble était transformé dans une souche de levure. Pour évaluer l'effet de chaque mutant sur la fonction p53, la capacité, du mutant p53 exprimé à transactiver la séquence de liaison de p53 (p53BS) dérivée du promoteur p21WAF1, était analysée en quantifiant la fluorescence d'EGFP émise. De plus l'activité transcriptionnelle à induire 7 autres gènes cibles (*MDM2*, *BAX*, *GADD45*, *14-3-3 σ* , *p53AIP1*, *NOXA* et *RRMB2* (*Ribonucleotide Reductase Regulatory TP53 inducible subunit M2B*)) était analysée en intégrant leur séquence de liaison dans

d'autres plasmides ayant Ds-Red comme gène rapporteur. Ce test a permis de classer l'effet de ces mutations en trois grandes classes : perte totale de l'activité transcriptionnelle, perte partielle et activité transcriptionnelle équivalente à la protéine sauvage.

Néanmoins, des limites concernant les tests préalablement décrits dans la levure concernent : l'utilisation de la mutagenèse dirigée sur l'ADNc de TP53 et donc l'exploration uniquement de la conséquence protéique des variations identifiées ; et la restriction de l'analyse aux mutations faux-sens pour le test de Kato. Ainsi les réarrangements génomiques de grande taille, les mutations aboutissant à un codon stop prématuré, et les mutations localisées au niveau intronique ou les mutations synonymes ne pourront pas être analysées. Pourtant, certaines mutations peuvent avoir des conséquences plus complexes, notamment sur l'épissage ou sur les niveaux d'ARN messager. De plus, même si l'activité de la protéine p53 humaine est maintenue dans la levure, celle-ci ne contient pas les nombreux cofacteurs humains nécessaires à la régulation de l'activité de p53. D'où le besoin de tests fonctionnels dans un modèle cellulaire humain afin d'évaluer cet effet probable des mutations de TP53 dans le contexte génétique des patients.

3- Essai fonctionnel ex vivo dans les lymphocytes immortalisés des patients LFS

Afin de tenter d'expliquer le gradient de sévérité clinique des mutations de TP53 sur le plan biologique et d'étudier dans le contexte génétique des patients LFS l'impact fonctionnel des différentes classes de mutations de TP53, nous avons développé dans notre laboratoire un essai fonctionnel simplifié ex vivo de la voie p53, à partir de lymphocytes immortalisés par le virus d'Epstein-Barr (EBV) exposés ou non à un stress génotoxique. Ces lymphocytes immortalisés ont été établis à partir du sang périphérique de patients LFS porteurs d'une mutation du gène TP53 à l'état hétérozygote et de 5 sujets contrôles de génotype TP53 sauvage (Zerdoumi et al., 2013). La protéine p53, facteur de transcription, est capable de transactiver de nombreux gènes cibles en réponse à un stress génotoxique. L'essai fonctionnel mis au point est basé sur l'activation de la voie p53 par traitement des lymphocytes du patient LFS et des témoins par la doxorubicine suivie par la mesure du taux d'induction de 6 gènes cibles de p53 par RT-QMPSF (Reverse Transcription – Quantitative Multiplex PCR of Short Fluorescent fragments). Ces 6 gènes choisis à partir des données de transcriptomes obtenues sont fortement et spécifiquement induits par p53 (KIAA0284, PODXL1, RRAD, GLS2, CABYR, TP53I3). Après superposition des profils de RT-QMPSF obtenus à partir des lymphocytes exposés et non exposés à la doxorubicine et ajustement sur 2 gènes contrôles (SF3A1 et TBP), le taux

d'induction de chaque gène cible est déterminé. La moyenne des taux d'induction des 6 gènes sélectionnés nous permet de définir un score de fonctionnalité de p53. Cet essai fonctionnel décrit dans la première partie des résultats, présente l'avantage de pouvoir appréhender l'effet dominant négatif de certaines mutations dans le contexte génétique du patient. Il représente un outil puissant pour classer les variations de signification biologique inconnue mais également pour appréhender la corrélation génotype - phénotype dans le LFS et pour détecter des patients présentant une inactivation de la voie p53, sans mutation identifiée.

Objectif 1

**Étude de l'impact des mutations constitutionnelles de *TP53*
sur la liaison à l'ADN de p53 et sur la réponse
transcriptionnelle aux lésions de l'ADN**

I- Contexte scientifique

Le syndrome de Li-Fraumeni représente l'une des formes héréditaires de cancers les plus graves avec des cancers de l'enfant et de l'adulte jeune. Il se caractérise par une forte hétérogénéité phénotypique tant au niveau du spectre tumoral que de l'âge de survenue des tumeurs. Ce syndrome résulte de mutations constitutionnelles hétérozygotes du gène suppresseur de tumeurs TP53, qui est le gène le plus fréquemment muté au niveau somatique dans les cancers sporadiques. A ce jour plus de 2000 familles évocatrices du LFS ont été analysées par l'unité de génétique moléculaire du CHU de Rouen et près de 300 familles avec mutation de TP53 ont été identifiées. Grâce à une étude nationale, portant sur 415 patients appartenant à 214 familles, présentant 133 altérations différentes de TP53, notre laboratoire a démontré qu'il existait chez les patients LFS un gradient de sévérité clinique des mutations constitutionnelles hétérozygotes de TP53, les mutations faux-sens avec un effet dominant négatif qui sont associées à un âge très précoce de survenue des tumeurs alors que les autres classes de mutations sont associées à une forme plus tardive (Bougeard et al., 2015). Grâce à l'essai fonctionnel développé par notre équipe sur les lymphocytes immortalisés de patients permettant d'évaluer très simplement l'intégrité de la réponse transcriptionnelle aux lésions médiée par p53, nous avons vérifié l'existence d'un gradient de sévérité biologique entre les mutations nulles de p53 et les mutations faux-sens associées à un effet dominant négatif qui sous-tend vraisemblablement le gradient de sévérité clinique observé chez les patients (Zerdoumi et al. 2013). Dans cette étude, nous avons d'abord voulu confirmer cette observation en effectuant le test fonctionnel de p53 dans des lymphocytes de 56 porteurs de mutation TP53 portant 35 altérations distinctes. Ensuite, nous avons voulu caractériser les mécanismes conduisant à la perte d'activité transcriptionnelle de p53 dans les lymphocytes de patients LFS porteurs de mutations hétérozygotes de TP53. Nous avons voulu vérifier l'hypothèse que le défaut observé dans la réponse transcriptionnelle liée à p53 chez les patients était lié à un défaut de fixation de la protéine p53 sur l'ADN. Pour vérifier cette hypothèse et l'hypothèse que moins de cibles de p53 sont reconnues dans les lymphocytes avec mutation à effet dominant-négatif que dans ceux avec mutation nulle, notre équipe a utilisé la technologie de séquençage haut-débit après immunoprécipitation de la chromatine ou ChIP-Seq. Cette technique permet d'identifier tous les sites de fixation de la protéine p53 à l'ADN en séquençant toutes les régions génomiques sur lesquelles se fixe p53, après exposition de lymphocytes humains à un stress génotoxique (la doxorubicine). Nous avons

réalisé l'expérience de ChIP-Seq sur des lignées témoins et sur des lignées mutées pour TP53 avec mutation faux-sens à l'état hétérozygote à effet dominant négatif et des lignées avec des mutations nulles (perte de fonction seule).

Afin de disposer d'un outil simple pour valider les résultats de cette technique pan-génomique, nous avons développé un test basé sur l'utilisation de la levure. Ce test indicateur de la fixation de p53 sur une séquence test est basé sur l'utilisation d'une souche de levure particulière qui est mutée pour le gène *ade2*. A partir de cette souche mutée pour *ade2* nous avons construit une souche qui exprime la séquence codante de p53, et en parallèle, une deuxième souche transformée avec le même vecteur mais sans p53. Un court amplicon est modélisé autour du fragment d'ADN à tester, identifié comme site de fixation de p53 par l'expérience de ChIP-Seq. Une première PCR est effectuée à partir de l'ADN génomique humain. Ensuite, à l'aide d'une deuxième PCR on rajoute à ce fragment les séquences qui vont être nécessaires à la recombinaison homologe dans la levure. Ces fragments de PCR sont transformés en même temps que le vecteur pRS313 linéarisé qui va exprimer le gène *ade2* seulement si le fragment de PCR inséré dans son promoteur par recombinaison homologe est reconnu par p53. Les levures transformées sont sélectionnées grâce au gène codant pour l'histidine présent sur ce plasmide. Après deux jours d'incubation des levures à 30 degrés, nous avons testé la présence d'un site de liaison de p53 à l'ADN par la simple lecture de la couleur des colonies. En effet la fixation de p53 à l'ADN active la transcription du gène *ade2* ce qui confère une couleur blanche aux colonies sinon les levures restent rouges. Ce test fonctionnel nous a permis de valider les résultats du ChIP-Seq et d'identifier de nouveaux gènes cibles de p53 (*GCNT3*, *LINC01480*, *PLXNB1* et *SRSF8*).

II- Résultats



Human Molecular Genetics, 2017, Vol. 26, No. 14 2591–2602

doi: 10.1093/hmg/ddx106

Advance Access Publication Date: 24 March 2017

Original Article

ORIGINAL ARTICLE

Germline TP53 mutations result into a constitutive defect of p53 DNA binding and transcriptional response to DNA damage

Yasmine Zerdoumi, Raphaël Lanos, Sabine Raad, Jean-Michel Flaman[†],
Gaëlle Bougeard, Thierry Frebourg* and Isabelle Tournier

Normandie University, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics,
F 76000, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

[†]To whom correspondence should be addressed at: Faculty of Medicine, Inserm U1245 and Rouen University Hospital, 22 Boulevard Gambetta, 76183 Rouen, France. Tel: +33 2 32 88 81 82; Fax: +33 2 32 88 80 80; Email: frebourg@chu-rouen.fr

Abstract

Li-Fraumeni Syndrome (LFS) results from heterozygous germline mutations of TP53, encoding a key transcriptional factor activated in response to DNA damage. We have recently shown, from a large LFS series, that dominant-negative missense mutations are the most clinically severe and, thanks to a new p53 functional assay in lymphocytes, that they alter the p53 transcriptional response to DNA damage more drastically than null mutations. In this study, we first confirmed this observation by performing the p53 functional assay in lymphocytes from 56 TP53 mutation carriers harbouring 35 distinct alterations. Then, to compare the impact of the different types of germline TP53 mutations on DNA binding, we performed chromatin immunoprecipitation-sequencing (ChIP-Seq) in lymphocytes exposed to doxorubicin. ChIP-Seq performed in wild-type TP53 control lymphocytes accurately mapped 1287 p53-binding sites. New p53-binding sites were validated using a functional assay in yeast. ChIP-Seq analysis of LFS lymphocytes carrying TP53 null mutations (p.P152Rfs*18 or complete deletion) or the low penetrant 'Brazilian' p.R337H mutation revealed a moderate decrease of p53-binding sites (949, 580 and 620, respectively) and of ChIP-Seq peak depths. In contrast, analysis of LFS lymphocytes with TP53 dominant-negative missense mutations p.R273H or p.R248W revealed only 310 and 143 p53-binding sites, respectively, and the depths of the corresponding peaks were drastically reduced. Altogether, our results show that TP53 mutation carriers exhibit a constitutive defect of the transcriptional response to DNA damage and that the clinical severity of TP53 dominant-negative missense mutations is explained by a massive and global alteration of p53 DNA binding.

Introduction

Li-Fraumeni syndrome (LFS; OMIM#151623) represents a paradigm in genetic predisposition to cancer considering the central role of p53 in the response to DNA damage and because our

perception of the syndrome, as defined by the presence of germline TP53 [MIM*191170, NM_000546.5] mutations, has considerably evolved since its original definition in 1969 by Frederick Li and Joseph Fraumeni (1–8). Indeed, a remarkable feature of the LFS is the heterogeneity of its clinical presentation, in terms of

[†]Present address: UMR Inserm U1052/CNRS 5286, Cancer Research Centre Lyon, France.

Received: February 21, 2017. Revised: February 21, 2017. Accepted: March 13, 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

2591

tumour type and age of tumour onset. In clinical practice, the presence of a germline *TP53* mutation should be considered in 4 different situations: (i) a proband with a LFS spectrum tumour (premenopausal breast cancer, soft tissue sarcoma, brain tumour, adrenocortical carcinoma (ACC)) before age 46 and at least one first or second-degree relative with a LFS tumour (except breast cancer, if the proband has breast cancer) before the age of 56 or with multiple tumours; (ii) a proband with multiple malignancies (except recurring breast cancer), two of which belonging to the LFS spectrum, the first being developed before 46 years; (iii) patients with rare early-onset tumours such as ACC, choroid plexus carcinoma (CPC) or embryonal anaplastic rhabdomyosarcoma, irrespective of the family history; and (iv) breast cancer before age 31 (1).

If germline *TP53* mutations are identified in familial aggregations of childhood and adult tumours, they can also be detected in patients and families who have developed only adult cancers and in patients without a familial history of cancer (1,9). With the exponential use, in cancer patients, of *TP53* tests performed with gene panels, it is likely that the percentage of germline *TP53* mutations detected in such patients or families will increase. This provides clinical evidence that germline *TP53* mutations display a heterogeneous penetrance. The diversity of the clinical presentations associated with germline *TP53* mutations suggests that it should be appropriate to include Li-Fraumeni syndrome within an expanded category designated 'TP53-related inherited cancers' characterized by a broader tumour spectrum and age of tumour onset.

All types of *TP53* alterations can be observed at the germline and at the somatic level. Nevertheless the *TP53* mutation spectrum is characterized by the predominance of missense mutations (above 65%), most of the mutations reported so far in LFS corresponding to dominant-negative mutations (1,10). LFS animal models and clinical data support the existence of genotype-phenotype correlations in LFS: (i) mouse models of LFS have shown the phenotypic severity of dominant-negative mutations, as compared with null mutations (11–14); (ii) from a large series of 415 *TP53* mutation carriers, we have recently reported that the mean age of tumour onset was statistically lower in carriers harbouring dominant-negative missense mutations than in carriers with loss of function mutations and that dominant-negative missense mutations represent the predominant germline alterations in carriers who developed childhood cancers, except ACC. In contrast, null mutations such as nonsense mutations, frameshift mutations or genomic rearrangements are predominantly detected in pedigrees characterized by cancers occurring in adulthood (1).

Thanks to the development of a new p53 functional assay in lymphocytes, based on the measurement of p53 transcriptional response to DNA damage, we provided a first molecular explanation of the dominant-negative missense mutation severity (15). Indeed, analysis of the p53 transcriptional response after doxorubicin exposure of lymphocytes derived from three *TP53* wild-type control subjects, 3 carriers of dominant-negative missense mutations and 3 carriers of null mutations, revealed that in lymphocytes with germline dominant-negative missense mutations, the number of induced genes and the level of p53 target gene induction were drastically reduced, as compared to controls and LFS lymphocytes with null mutations. We subsequently developed a simpler version of the assay, based on the transcriptional induction of a limited number of p53 target genes in lymphocytes exposed to DNA damaging agents. This assay, performed in wild-type *TP53* lymphocytes, can be used to determine and quantify the genotoxic effect of chemical or physical agents, p53

being a universal sensor of genotoxic stress (16). It can alternatively be used to determine the functionality of p53 protein in LFS patients harbouring germline *TP53* mutations.

In this study, to deeply characterize the biological impact of the different classes of germline *TP53* mutations, we first analysed, using a modified version of this functional assay, the effect of 35 distinct germline alterations on p53 transcriptional response to DNA damage and then characterized the consequences of germline *TP53* mutations on DNA binding, using chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq).

Results

p53 transcriptional response to DNA damage in lymphocytes constitutes an endophenotype of germline *TP53* mutation severity

To confirm the results obtained in our previous study performed on a limited number of samples and showing that dominant-negative missense mutations had a more drastic impact on p53 transcriptional response compared to null mutations (15), we selected 35 EBV-immortalized lymphocytes from 5 control subjects with wild-type *TP53* genotype, 14 LFS patients with *TP53* null mutations and 16 LFS patients with missense mutations previously classified as dominant-negative mutations according to the IARC (International Agency for Research on Cancer) *TP53* database (<http://p53.iarc.fr/>; date last accessed March 22, 2017) (17) (Table 1). We analysed the p53 transcriptional response to DNA damage in these EBV-immortalized lymphocytes, using our simplified p53 functional assay, based on the transcriptional induction of six p53 target genes after exposure of lymphocytes to doxorubicin, a genotoxic agent inducing DNA double-strand breaks, and on the subsequent comparison of p53 target gene expression levels between doxorubicin-treated and non-treated cells (Supplementary Material, Fig. S1). We checked that p53 was efficiently activated following exposure to doxorubicin, by monitoring p53 protein stabilization (Supplementary Material, Fig. S2). In this assay, the p53 functional activity is expressed as an arbitrary functionality score corresponding to the average of the six p53 target gene induction levels. As shown in Figure 1, in control wild-type *TP53* lymphocytes, we obtained a mean p53 functionality score of 11.6 ± 1.04 , indicating a strong induction of p53 transcriptional activity in response to DNA damage. In LFS lymphocytes with null mutations, we observed a 50% decrease of the mean score (6.7 ± 0.4) (Fig. 1A) illustrating the p53 haplodeficiency (Table 1). In LFS lymphocytes with dominant-negative missense mutations, we detected a drastic reduction of the p53 transcriptional activity (mean score = 3.0 ± 0.2) (Fig. 1A). The mean scores were statistically different between control wild-type *TP53* lymphocytes and those with null mutations (P -value < 0.0001, Student *t*-test) or with dominant-negative missense mutations (P -value < 0.0001, Student *t*-test) (Fig. 1A). As shown in Table 1, we obtained similar results for EBV-immortalized lymphocytes derived from different individuals harbouring the same germline *TP53* mutation, indicating the reproducibility of the assay. These results therefore confirm our previous results showing that germline *TP53* mutations detected in LFS patients result in a constitutive reduction of p53 transcriptional activity and that dominant-negative missense mutations have a more severe impact on this transcriptional activity than null mutations.

We then extended this functional analysis to 26 patients suspected to present with LFS and harbouring 15 additional different missense mutations. According to the IARC *TP53* database (17), 3 had no detectable dominant-negative effect, and for

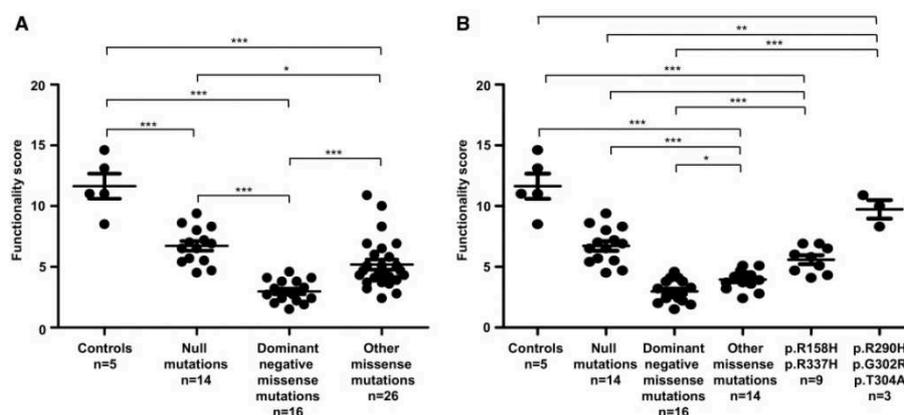


Figure 1. Comparative analysis of p53 transcriptional activity in wild-type and mutant TP53 EBV-immortalized lymphocytes exposed to doxorubicin. (A) Dot plot depicting the comparison of p53 functionality scores obtained using p53 functional assay, in lymphocytes from 5 wild-type TP53 controls and from 56 TP53 mutation carriers. For each type of mutations 'n' indicates the number of carriers analysed. (B) Subdivision of the 26 patients harbouring 15 different missense mutations into 3 groups: (i) mutations classified according to the IARC database, as supertrans or functional (p.R290H, p.G302R, p.T304A) (ii) p.R158H and p.R337H mutations, and (iii) other missense mutations (p.E11Q, p.G108D, p.T125M, p.R175G, p.H193D, p.H193P, p.V216M, p.C229R, p.I254T, p.R282P). Scores correspond to the average of two independent experiments; the error bars represent mean \pm standard error of the mean (SEM). Statistical significance of the comparisons was evaluated using the unpaired t-test and represented by an asterisk (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

12 mutations the dominant-negative activity had not been assessed (Table 1). For these 15 mutations we observed, as shown in Figure 1A and Table 1, a dispersion of the p53 functionality scores, indicating that these mutations have a heterogeneous effect on p53 transcriptional activity, which led us to subdivide these mutations into 3 groups (Fig. 1B). The first group corresponded to 3 mutations (p.R290H, p.G302R, p.T304A) previously classified as 'supertrans or functional' in a yeast-based functional assay (18). The scores for these mutations (mean score = 9.7 ± 0.8) were very close to the control lymphocyte scores (Fig. 1B, Table 1), confirming the absence of the deleterious effect of these mutations on the p53 transcriptional activity. When we looked at the TP53 missense mutations p.R158H and p.R337H associated with adrenocortical carcinoma or adult tumours (1,19), we obtained scores close to the ones of the null mutations (mean score = 5.6 ± 0.4) (Fig. 1B, Table 1). Finally, when we analysed the other missense mutations (p.E11Q, p.G108D, p.T125M, p.R175G, p.H193D, p.H193P, p.V216M, p.C229R, p.I254T, p.R282P), we observed a drastic alteration of the p53 transcriptional activity (mean score = 3.8 ± 0.2), similar to that observed with well characterized dominant-negative missense mutations suggesting that most of these mutations are also *bona fide* dominant-negative mutations (Fig. 1B, Table 1). These results indicate that the impact of the mutations on p53 transcriptional response to DNA damage in LFS lymphocytes can be considered as an endophenotype of the clinical severity of germline TP53 mutations.

Genome-wide mapping of p53-binding sites in control lymphocytes exposed to DNA damage

In order to evaluate the impact of germline TP53 mutations on p53 DNA binding, we did a genome-wide mapping of p53-binding sites in response to DNA damage, by performing ChIP-Seq in control EBV-immortalized lymphocytes exposed to

doxorubicin. We used EBV-immortalized lymphocytes derived from two TP53 wild-type individuals in order to limit a potential bias due to inter-individual variability and we performed each experiment in duplicate. In control subject 1, we identified 2969 and 2249 peaks (FDR = 0.005) in replicates R1 and R2, respectively, with 1963 peaks in common between the two replicates. In control subject 2, 1684 and 1986 peaks were identified in replicates R1 and R2, with 1442 common peaks (Figs 2A and 3A). As expected, few peaks were detected when the same experiment was performed without doxorubicin treatment (Supplementary Material, Fig. S3). In order to identify the most relevant p53-binding sites, we selected peaks detected in common between replicates performed on lymphocytes from the two controls, which allowed us to identify 1287 high-confidence peaks (FDR = 0.005) (Fig. 2A, Supplementary Material, Table S1), among which 65% (842/1287) and 26% (337/1287) were located within 50 kb and 10 kb from transcription start sites (TSS), respectively. Moreover, 90% (1158/1287) of these high-confidence ChIP-Seq peaks harbour the previously characterized p53-binding motif (20) (Supplementary Material, Table S1).

In order to identify, among these genes containing p53-binding sites, those who are differentially expressed upon p53 activation, we completed the transcriptomic analysis previously performed on lymphocytes derived from the control 1 and exposed to doxorubicin (15), by a new transcriptomic analysis performed on lymphocytes derived from control 2. Global gene expression analysis allowed us to identify in exposed control lymphocytes a core of 913 differentially expressed genes, including 532 up and 381 down-regulated genes (P -value < 0.005) (Supplementary Material, Table S2). Comparison of ChIP-Seq data with gene expression profiles showed that 12% (143/1208) of the genes containing a p53-binding site were differentially expressed (127 up-regulated and 16 down-regulated genes) upon doxorubicin treatment (Fig. 2B and C). These 143 genes include 40 previously known p53 target genes (21,22), such as CDKN1A [MIM*116899, NM_000389] and MDM2 and 103 new

Table 1. Functionality scores in control subjects and TP53 mutation carriers

Individual (gender ^a , age ^b)	Family	Phenotype ^c (age ^d)	Germline TP53 mutation ^e	p53 mRNA expression ^f	Score ^g
Controls					
Control 1 (M, 34)	F1	No tumour	None	79	11
Control 2 (F, 37)	F2	No tumour	None	100	13.1
Control 3 (F, 30)	F3	No tumour	None	100	14.6
Control 4 (M, 29)	F4	No tumour	None	71	11
Control 5 (M, 28)	F5	No tumour	None	87	8.5
Missense mutations with dominant negative effect (DNE)^h					
Patient 1 (F, 29)	F155	Bilat. ACC (1 + 2), SC (28)	c.524G>A, p.R175H	93	2.9
Patient 2 (F, 2)	F65	anRMS (1)	c.524G>A, p.R175H	86	4.1
Patient 3 (F, 14)	F24	GBM (14), OS (16)	c.535C>T, p.H179Y	78	3.8
Patient 4 (M, 1)	F225	CPC (0.3)	c.638G>A, p.R213Q	98	4.1
Patient 5 (F, 23)	F185	HD (13), BC (21), BS (21)	c.659A>G, p.Y220C	91	3.3
Patient 6 (F, 23)	F161	BC (22)	c.733G>A, p.G245S	93	2.4
Patient 7 (M, 52)	F241	No tumour (52)	c.733G>A, p.G245S	85	3.8
Patient 8 (F, 26)	F254	Bilat. BC (25)	c.736A>G, p.M246V	90	2.7
Patient 9 (F, 29)	F267	OS (16), BC (28)	c.743G>A, p.R248Q	76	2.2
Patient 10 (F, 37)	F212	BC (25)	c.742C>T, p.R248W	70	1.5
Patient 11 (F, 35)	F141	STS (19), Bilat. BC (29 + 34)	c.742C>T, p.R248W	73	3.2
Patient 12 (M, 22)	F141	Bilat. tes. GCT (17 + 19), BCC (19)	c.818G>A, p.R273H	104	1.9
Patient 13 (F, 3)	F209	STS (2)	c.818G>A, p.R273H	88	2.4
Patient 14 (F, 9)	F145	No tumour (10)	c.844C>T, p.R282W	104	2
Patient 15 (F, 35)	F79	No tumour (35)	c.844C>T, p.R282W	116	2.7
Patient 16 (F, 9)	F79	No tumour (9)	c.844C>T, p.R282W	76	4.6
Null mutations					
Patient 17 (F, 20)	F26	No tumour (20)	c.(?-202)_(?1207?) del, p.0(comp del) ⁱ	57	7.2
Patient 18 (F, 66)	F26	BC (46), STS (49)	c.(?-202)_(?1207?) del, p.0(comp del) ⁱ	42	6.5
Patient 19 (M, 16)	F35	No tumour (16)	c.(?-202)_(?29+1_-28-1)del, p.0 (prom-ex1 del) ^j	53	5.4
Patient 20 (M, 60)	F167	STS (59)	c.(?-202)_(?29+1_-28-1)del, p.0 (prom-ex1 del) ^j	60	8
Patient 21 (M, 28)	F56	No tumour (28)	c.216dup, p.V73Rfs*76	51	4.5
Patient 22 (F, 28)	F203	BC (27)	c.216dup, p.V73Rfs*76	44	6.5
Patient 23 (F, 31)	F208	BC (28)	c.323_329dup, p.L1111Ffs*40	59	6.9
Patient 24 (F, 29)	F211	Bilat. BC (26 + 29)	c.455del, p.P152Rfs*18	63	4.7
Patient 25 (M, 12)	F223	OS (11)	c.491_494del, p.K164Sfs*5	50	8.6
Patient 26 (M, 40)	F223	Tes. GCT (26), BCC (40)	c.491_494del, p.K164Sfs*5	68	8.3
Patient 27 (F, 27)	F186	Bilat. BC (25)	c.632_641del, p.T211Ifs*33	50	5.7
Patient 28 (F, 29)	F146	No tumour (29)	c.690del, p.T231Pfs*16	52	9.4
Patient 29 (F, 12)	F162	ACC (11)	c.820del, p.V274Ffs*71	53	7
Patient 30 (F, 35)	F28	No tumour (35)	c.673-2A>G, p.?	42	5.5
Missense mutations with no detectable DNE or whose dominant activity had not been assessed					
Patient 31 (M, 3)	F324	NB (2)	c.31G>C, p.E11Q	120	3
Patient 32 (F, 58)	F150	No tumour (58)	c.323G>A, p.G108D	81	4.2
Patient 33 (F, 2)	F204	ACC (0.9)	c.374C>T, p.T125M	92	5.1
Patient 34 (F, 16)	F149	ACC (0.5)	c.473G>A, p.R158H ^k	83	4.1
Patient 35 (M, 3)	F188	CPC (2)	c.473G>A, p.R158H ^k	93	4.3
Patient 36 (F, 18)	F41	No tumour (18)	c.473G>A, p.R158H ^k	89	4.7
Patient 37 (M, 18)	F41	No tumour (18)	c.473G>A, p.R158H ^k	72	5.1
Patient 38 (F, 18)	F166	No tumour (18)	c.473G>A, p.R158H ^k	83	6.9
Patient 39 (F, 24)	F2	OS (18), bilat. BC (27 + 29), LC (45)	c.523C>G, p.R175G ^k	101	3.2
Patient 40 (F, 33)	F206	STS (24), BC (26)	c.577C>G, p.H193D	91	3.7
Patient 41 (M, 57)	F181	STS (49), LC (54), MM (55)	c.578A>C, p.H193P	109	3.7
Patient 42 (M, 30)	F144	ARMS (3), STS (17)	c.646G>A, p.V216M	94	4.3
Patient 43 (M, 34)	F179	No tumour (37)	c.646G>A, p.V216M	82	4.3
Patient 44 (F, 37)	F238	BC (34)	c.685T>C, p.C229R	82	4.3
Patient 45 (F, 36)	F198	No tumour (36)	c.761T>C, p.I254T	72	2.4
Patient 46 (F, 29)	F198	No tumour (29)	c.761T>C, p.I254T	72	2.8
Patient 47 (M, 1)	F198	ACC (0.9)	c.761T>C, p.I254T	104	3.6
Patient 48 (F, 32)	F198	No tumour (32)	c.761T>C, p.I254T	74	3.9
Patient 49 (F, 24)	F148	CPC (17), BS (24)	c.845G>C, p.R282P	105	4.7

(continued)

Table 1. (continued)

Individual (gender ^a , age ^b)	Family	Phenotype ^c (age ^d)	Germline TP53 mutation ^e	p53 mRNA expression ^f	Score ^g
Patient 50 (F, 67)	F160	BC (47)	c.869G>A, p.R290H ^h	85	10.9
Patient 51 (F, 44)	F142	BC (41)	c.904G>A, p.G302R	71	10
Patient 52 (F, 33)	F77	Ov. GCT+STS (28)	c.910A>G, p.T304A	78	8.3
Patient 53 (M, 2)	F195	ACC (2)	c.1010G>A, p.R337H	81	5.8
Patient 54 (M, 38)	F259	No tumour (38)	c.1010G>A, p.R337H	85	6
Patient 55 (F, 1)	F216	ACC (0.6)	c.1010G>A, p.R337H	81	6.5
Patient 56 (M, 4)	F265	ACC (3)	c.1010G>A, p.R337H	101	6.9

^aM, male; F, female.

^bAge (in years) at blood sampling for EBV-immortalization of lymphocytes.

^cACC, adrenocortical carcinoma; anRMS, anaplastic rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma; BC, breast cancer; BCC, basal cell carcinoma; Bilat, bilateral; BS, breast sarcoma; CPC, choroid plexus carcinoma; GCT, germ cell tumour; GBM, glioblastoma multiforme; HD, Hodgkin's disease; LC, lung cancer; MM, malignant melanoma; NB, neuroblastoma; OS, osteosarcoma; Ov, ovary; SC, skin cancer; STS, soft tissue sarcoma; Tes, testis.

^dAge (in year) corresponds either to the age at tumour diagnosis or to the age of the unaffected individual.

^ecDNA numbering with the first nucleotide corresponding to the A of the ATG translation initiation codon in the reference sequence (GenBank RefSeq-file accession number NM_000546.5), according to the Human Genome Variation Society guidelines (www.hgvs.org/mutnomen; date last accessed March 22, 2017). The initiation codon is numbered as codon 1 (accession number NP_000537.3).

^fIndicated values correspond to p53 mRNA level; data are expressed as percentage of the value observed in control 2.

^gIndicated values correspond to the mean of the average fold-changes of target genes expression obtained from two independent p53 functional assays.

^hAccording to the IARC TP53 database.

ⁱComplete deletion.

^jPromoter and exon 1 deletion.

^kWith no detectable DNE according to the IARC TP53 database.

putative p53 target genes (Supplementary Material, Table S2). As expected, functional clustering of gene annotations, using the DAVID algorithm, revealed an enrichment of genes involved in biological pathways known to be regulated by p53 in response to DNA damage (Supplementary Material, Table S2).

Functional validation of new putative p53 target genes using a p53-binding assay in yeast

To confirm the functional relevance of the p53-binding sites identified by ChIP-Seq within new putative p53 target genes, we constructed 2 yeast strains, one expressing human wild-type p53 (YPH500-TP53) and the other containing the empty vector (YPH500-Empty). p53-binding sites, detected by ChIP-Seq, were PCR-amplified and cloned by homologous recombination into the gapped plasmid pRS313-cyc-ADE2 in the minimal promoter upstream of the ADE2 open reading frame and transformed into the YPH500-TP53 and YPH500-Empty yeast strains. In this assay, the ADE2 gene expression is strictly dependent on p53-binding to the cloned fragment. Colonies of YPH500 strains cultured on media containing a limited amount of adenine are spontaneously red, whereas colonies expressing ADE2 turn white (Supplementary Material, Fig. S4).

From the ChIP-Seq data, we selected 4 binding sites in genes not known as p53 target genes and absent from previous ChIP-Seq studies (GCNT3, LINC01480, PLXNB1 and SRSF8) (23) and tested them in the yeast binding assay. As a positive control, we used the p53-binding site within MDM2 and, as a negative control, a sequence from the USH1G gene [MIM*607696, NM_173477] without any signal detected in the p53 ChIP-Seq experiments. All 4 putative p53-binding sites generated white colonies when transformed into YPH500-TP53 strain and red colonies when transformed into the YPH500-Empty yeast strain, confirming that they indeed correspond to functional p53-binding sites (Supplementary Material, Fig. S5). We then performed quantitative RT-PCR analysis to confirm the results obtained in the transcriptomic analyses for these genes. In accordance with the

transcriptomic data, all the genes examined were differentially expressed in control lymphocytes exposed to doxorubicin except for SRSF8 (Supplementary Material, Fig. S5, Supplementary Material, Table S2). In lymphocytes harbouring the TP53 p.R248W mutation, the up-regulation of these genes was significantly reduced, demonstrating the specificity of the regulation by p53 (Supplementary Materials, Fig. S5, Table S2).

Drastic effect of TP53 dominant-negative missense mutations on p53 DNA binding

In order to study the functional consequences of germline TP53 mutations on p53 DNA binding in presence of DNA damage, we performed a ChIP-Seq analysis of EBV-immortalized lymphocytes from LFS patients harbouring heterozygous TP53 mutations, after exposure to doxorubicin. For this study, we selected (i) two null mutations, i.e. a frameshift mutation (p.P152Rfs*18) and a complete deletion, both resulting into p53 haploinsufficiency, (ii) two canonical missense mutations, classified as dominant-negative (p.R248W, p.R273H), according to the IARC TP53 database and (iii) the p.R337H mutation, a low penetrance TP53 mutation associated with adrenocortical carcinomas or adulthood tumours (Table 1). As shown in Figure 3A, the total number of ChIP-Seq peaks detected was reduced approximately by half in LFS lymphocytes harbouring the p.P152Rfs*18 mutation or the complete TP53 deletion compared to the two controls. In LFS lymphocytes harbouring dominant-negative missense mutations, a massive reduction was observed (Fig. 3A). Interestingly, the number of peaks obtained for the p.R337H missense mutation was closer to those obtained for null mutations than those obtained for dominant-negative mutations (Fig. 3A). We then compared the peak scores of the p53-binding sites, detected in wild-type and mutant lymphocytes. We identified 845, 562, 532, 239 and 139 peaks in common between control lymphocytes and LFS lymphocytes harbouring p.P152Rfs*18, complete deletion, p.R337H, p.R273H and p.R248W mutations, respectively (Fig. 3B). In each case, the peak score in mutant

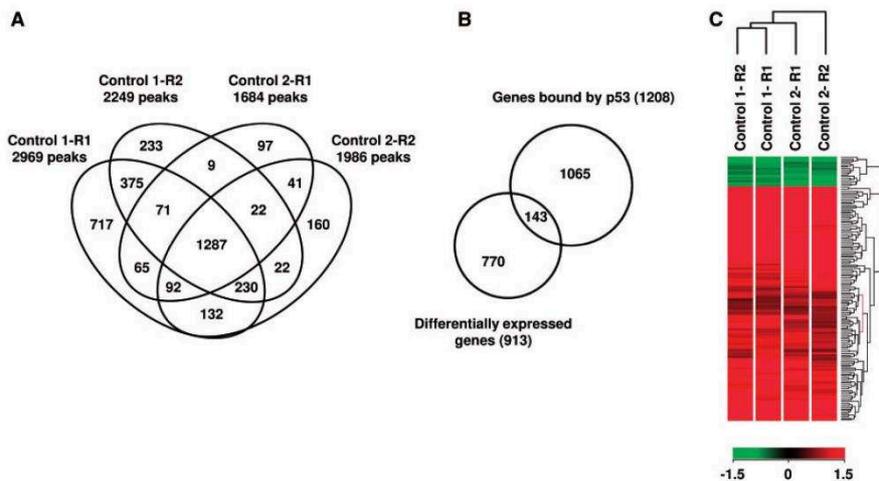


Figure 2. Identification by ChIP-Seq analysis of p53-binding sites in control EBV-immortalized lymphocytes exposed to doxorubicin. (A) Venn diagram illustrating the overlaps of peaks detected in common between the 2 replicates (R1 and R2) performed in 2 controls (control 1 and control 2). The numbers indicate the number of binding sites recognized by wild-type p53. (B) Overlap between genes containing p53-binding sites and genes differentially expressed upon doxorubicin treatment (note that the 1287 high-confidence peaks correspond to 1208 different genes). (C) Heat map, obtained in lymphocytes from two wild-type TP53 controls, showing differential expression upon doxorubicin treatment of 143 genes containing a p53-binding site. For each control, replicates are indicated (R1 and R2). Red color indicates up-regulated genes and green color down regulated genes.

lymphocytes was significantly reduced ($P < 0.0001$ unpaired *t*-test), as compared with control lymphocytes (Fig. 3B). As shown for the 123 peaks detected in common in wild-type and mutant lymphocytes, (Fig. 3C) and as illustrated by the visualization of peaks at the *CDKN1A* locus (Fig. 3D), a well-known p53 target gene, we observed a massive reduction of the peak scores in LFS lymphocytes harbouring dominant-negative missense mutations, as compared to control lymphocytes or LFS lymphocytes harbouring null mutations. Again, the results obtained for the p.R337H mutation were similar to those obtained with null mutations (Fig. 3C and D).

No evidence for gain of function in lymphocytes with missense mutations

In addition to loss of function and dominant-negative effect, some mutant p53 can exert additional oncogenic activity, by gain-of-function (GOF) mechanism. Because some studies have shown that some dominant-negative missense mutations, such as the p.R248W mutation, may result in GOF by modulating expression of target genes involved in oncogenesis (24,25), we explored thanks to the ChIP-Seq data, the putative GOF activity of the p.R248W mutant in lymphocytes not exposed to doxorubicin. As expected in this basal condition, only 17 peaks were detected (Supplementary Materials, Fig. S3A, Table S1). Among these peaks, none was specific to the mutant. These results indicate that in non-malignant cells, such as lymphocytes, there is no evidence that missense mutations induce GOF.

Discussion

In this study, we explored the impact of germline TP53 mutations on the transcriptional activity of the protein in response

to DNA damage. Since 1992, it has been established that the tumour suppressor activity of p53 relies mostly on its ability, when DNA damage occur, to upregulate the transcription of numerous target genes involved in many pathways including cell cycle arrest, DNA repair, apoptosis and metabolism. The use of the simple p53 functional assay that we developed in lymphocytes, based on the semi-quantitative measurement of the induction of six p53 target genes, confirms that in TP53 mutation carriers there is a constitutive defect in the transcriptional response to DNA damage, detectable in most of the cases. Indeed, the results of the p53 functional assay obtained for 30 carriers with TP53 null or dominant-negative missense mutations clearly showed, except for 4 carriers harbouring null mutations, that the transcriptional activity was altered as compared to controls (Table 1). It should be highlighted that the main advantage of this assay developed in lymphocytes, although it requires immortalization of lymphocytes by EBV, is that it allows one to assess the biological impact of heterozygous mutations in the genetic context of the patients, in contrast to *in vitro* and *ex vivo* assays, such as those developed in yeast (18,26,27). We have compared the p53 transcriptional response to DNA damage between peripheral blood mononuclear cells and EBV-immortalized lymphocytes, and found that the transcriptional response to doxorubicin was similar (data not shown). Moreover, the absence of p53 stabilization in unexposed EBV-immortalized lymphocytes from 5 controls (Supplementary Material, Fig. S2) confirms that EBV immortalization does not introduce genetic changes activating p53 pathway.

In TP53 mutation carriers, this constitutive alteration of the transcriptional response to DNA damage, detectable in non-tumour cells, constitutes an endophenotype before the occurrence of cancer and confirms that germline TP53 mutations represent a genetic permissive context facilitating the

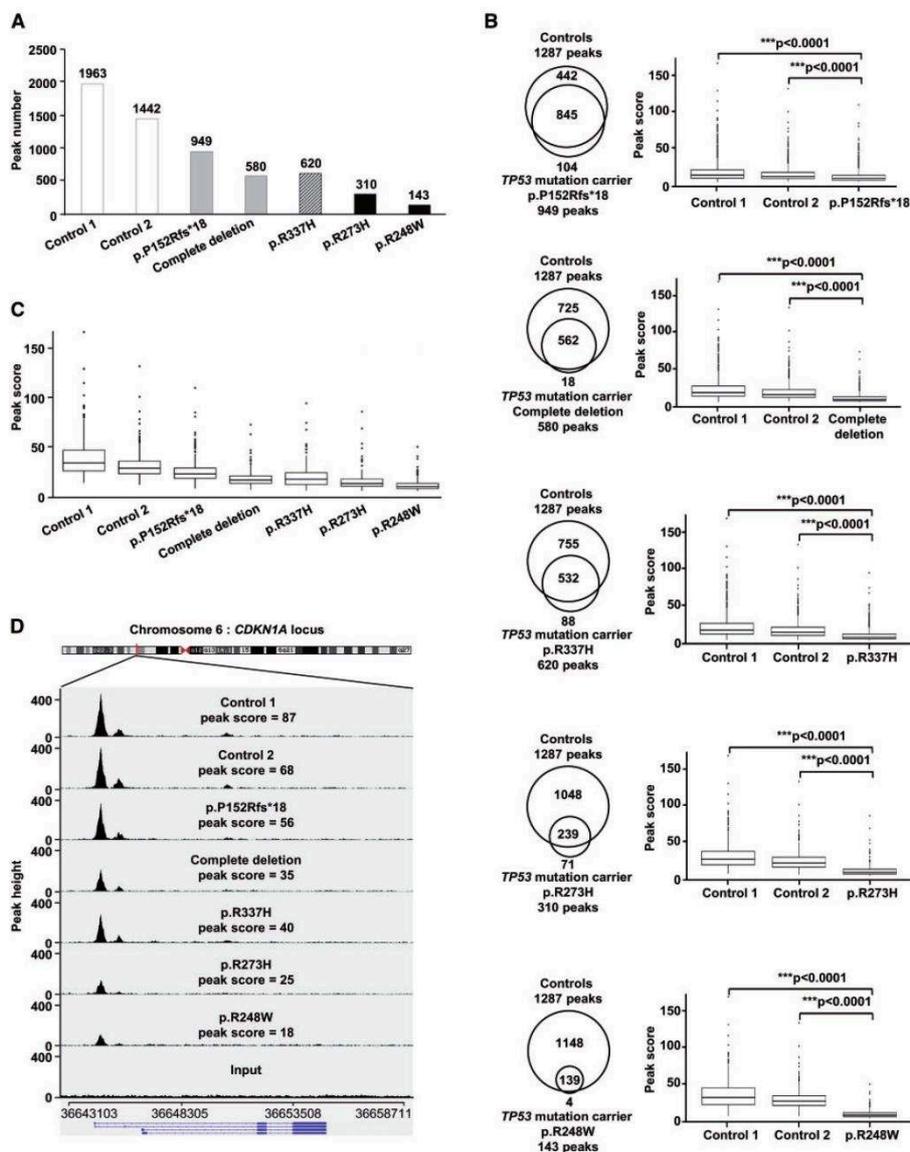


Figure 3. Impact of germline *TP53* mutations on DNA-binding in lymphocytes exposed to doxorubicin, revealed by ChIP-Seq analysis. (A) Histograms presenting the number of peaks detected by ChIP-Seq performed on lymphocytes derived from two wild-type *TP53* controls (control 1 and control 2) (open bars), two *TP53* mutation carriers harbouring null mutations (p.P152Rfs*18 and complete deletion) (grey bars), one carrier harbouring the p.R337H mutation (hatched bars), and two carriers harbouring dominant-negative missense mutations (p.R273H and p.R248W) (black bars); (B) Venn diagrams showing, for each *TP53* mutation, the overlaps of ChIP-Seq peaks detected in common between control lymphocytes and mutant lymphocytes. The ends of the boxes correspond to 25 and 75% quartiles, the median is marked by a horizontal line inside the box, the lower whiskers on the box plots represent the smallest peak score $\geq 25\%$ quartile $- 1.5$ (IQR) (inter-quartile range), and the upper whisker represent the largest peak score $\leq 75\%$ quartile $+ 1.5$ (IQR). Peak scores beyond the end of the whiskers are the outliers and plotted as individual points. Unpaired t-tests were performed to compute significance. (C) Box plot analysis of the average peak scores of the 123 peaks detected in common between all control and *TP53* mutant lymphocytes. (D) IGV visualization of p53 ChIP-Seq results at the *CDKN1A* locus in control lymphocytes, lymphocytes harbouring null mutations, the p.R337H mutation or dominant negative missense mutations. The average ChIP-Seq peak scores, detected at the *CDKN1A* locus, in both replicates of ChIP-Seq performed on lymphocytes from the two controls (control 1 and control 2) and each the *TP53* mutation carriers (p.P152Rfs*18, complete deletion, p.R337H, p.R273H and p.R248W) are indicated.

malignant transformation of cells in which DNA damage has occurred. This suggests that the core LFS tumour spectrum is probably explained by the differential sensitivities of tissues to oncogenic stresses and that the tumour spectrum associated with germline TP53 mutations is much broader than initially considered. This also represents an additional argument supporting that genotoxic treatments, such as radiotherapy, contribute to secondary tumours occurring in LFS patients and therefore should be avoided, if possible. Thus, LFS represents a rare example of autosomal dominant predisposition to cancers, in which a biological endophenotype can be detected in mutation carriers. As shown in Table 1 and Figure 1, the results obtained with the functional assays confirm the difference in the biological severity of dominant-negative missense mutations versus the other types of missense mutations and the null mutations. This assay should help the classification of TP53 mutations and the interpretation of variants of unknown significance. We anticipate that this challenge will increase with the detection of such germline TP53 variants in patients with atypical clinical presentations, performed in the context of next-generation sequencing (NGS) cancer gene panels. A normal score obtained with this p53 functional assay will not exclude that the variant is deleterious, since the assay explores only the p53-mediated transcriptional response to damage in lymphocytes and, furthermore, we observed for some null mutations a score close to that obtained in controls (Table 1). In contrast, an abnormal score should constitute a strong argument supporting the deleterious effect of the variant.

Several studies have previously analysed p53 DNA-binding in cells exposed to genotoxic or non-genotoxic agents, using ChIP-Seq. Most of these studies have been performed in established cancer cell lines, such as HCT116 or U2OS derived from human colon carcinoma and osteosarcoma, respectively (28–31). Our study provides new insights concerning p53 binding to DNA in non-cancer cells, when DNA damage occurs. First, ChIP-Seq performed in replicates on control lymphocytes allowed us to detect 1287 high-confidence p53-binding sites in the presence of DNA damage, and 90% of these high-confidence peaks harboured a p53-binding motif (20). We showed, by gene expression analyses and yeast assays, that a fraction of these p53-binding sites corresponds to real p53 target genes. The comparison of p53 gene occupancy and gene expression profiles in control lymphocytes exposed to doxorubicin reveals that 12% of the genes bound by p53 were differentially expressed upon doxorubicin treatment (Fig. 2B and C). The 770 differentially expressed genes not bound by p53 could represent genes regulated by a distal p53-binding site or genes indirectly regulated by p53 through other transcription factors. On the other hand, we showed that 1065 genes specifically bound by p53 were not differentially expressed in response to doxorubicin. This lack of induction could either reflect a bias in the annotation of the peaks which were arbitrary associated with the nearest transcription start site since the bound elements could represent distal enhancers for another gene; it could also be explained by the lack of a cell-specific or context-specific cofactor needed for p53 transcriptional activity. Indeed, several p53 ChIP-Seq studies have shown that in different cells treated with different genotoxic (doxorubicin, 5-Fluorouracil) or non-genotoxic agents (RITA, Nutlin-3a), the p53-transcriptional response is different despite a mostly similar binding profile (30–32). This observation suggests that changes in gene expression induced by p53 require other factors, such as cofactors recruited in a cell type-specific and stress response-specific manner, indicating that p53 binding is necessary, but not sufficient for transactivation.

Alternatively, p53 chromatin occupancy in response to DNA lesions without transactivation may reflect other p53 roles in chromatin integrity (33).

In agreement with recent studies (34,35), our data show that, in non-damaged cells, p53 is already bound to a small proportion of its genomic targets, and 100% of these p53-binding sites remain when DNA damage occurs (Supplementary Material, Fig. S3B), suggesting that this constitutive fixation of p53 to target genes allows a rapid response to genotoxic stress. As shown by the results of the ChIP-Seq performed on control lymphocytes exposed to doxorubicin, induction of DNA damage results not only in an increase of peak scores (a measure of p53-binding strength), at sites detectable in basal conditions, but also into a massive increase of the total number of p53-binding sites (Supplementary Material, Table 1, Supplementary Material, Fig. S3). Our results contrast with those obtained in cancer cell lines exposed to doxorubicin (30). Indeed, in that study, 90% of the p53-binding sites detected in untreated cells did not contain the p53-binding motif, and in treated cells many of these sites were not bound by p53. The differences between both studies might be explained by the numerous genetic and epigenetic modifications occurring in cancer cell lines (36), which deeply modify the landscape of p53-binding sites (28,29). Indeed, most cancer cell lines harbour either a TP53 mutated gene or an alteration in the p53 signalling pathway (37,38).

In lymphocytes from LFS patients harbouring null mutations, the p53 functionality scores obtained with the p53 functional assay, the number and scores of ChIP-Seq peaks were reduced by half in average, as compared to control lymphocytes (Figs 1 and 3). These results are those expected for haploinsufficiency due to a 50% reduction of the mRNA expression level (Table 1). This observation suggests that the DNA binding of wild-type p53 protein, encoded by the wild-type allele allows maintaining a partial p53 transcriptional response to DNA damage. This probably explains why TP53 null mutations are associated with later tumour-onset and why in tumours with such mutations, LOH affecting the wild-type allele is more commonly observed than in tumours with other types of TP53 mutations (1,39).

The drastic effect of dominant-negative missense mutations on p53 transcriptional activity and DNA binding in response to DNA damage (Figs 1A, 3A–D) is probably explained by the ability of the p53 dominant-negative mutants to disrupt the wild-type p53 function by forming inactive heterotetramers. Indeed, it has been suggested that p53 dimers are formed co-translationally, resulting in either mutated or wild-type homo-dimers, whereas tetramers are formed post-translationally by dimerization of the existing dimers (40). According to this model, the mutant p53 protein would inactivate 75% of p53 tetramers (41). This is consistent with our data showing a reduction close to 75% of the functionality score and the number of detected peaks in LFS lymphocytes with dominant-negative missense mutations in comparison to those observed in control lymphocytes. Thus, we provide a biological explanation of the clinical severity of the dominant-negative missense mutations which drastically alter the p53 transcriptional activity in response to DNA damage by abrogating binding to DNA.

The observation of germline TP53 null mutations strongly supports that the *primum movens* of LFS is the loss of p53 function. Therefore, the predominance of missense mutations and, in particular of those with dominant-negative activities, constitutes an apparent paradox. The drastic effect of these mutations on DNA binding and transcriptional response to DNA, that we report here, probably may explain only in part their

predominance and clinical severity. The second part of the explanation probably corresponds to the gain of multiple neomorphic properties, TP53 missense mutations being able to transform the tumour suppressor protein into an oncogenic protein, providing a selective advantage during cancer progression. As highlighted by numerous studies, GOF includes alteration of chromatin structures, genomic instability, inhibition of DNA repair, interaction with transcription factors, aberrant upregulation of genes and subversion of molecular partners of wild-type p53 (42,43). A recent study, also based on ChIP-Seq but performed in cancer cell lines, has reported that missense mutations with GOF cause aberrant DNA binding and upregulation of genes involved in chromatin conformation (25). However, our analyses performed in lymphocytes with missense mutations did not provide any evidence for aberrant transcription or DNA binding. As previously shown by transcriptomic analyses performed in non-cancer tissues from p53^{-/-} and p53 mutant/mutant mice, which did not reveal any significant changes (44), our study confirms that GOF is not observed in non-malignant cells and suggests that GOF occurs when cells are transformed.

The clinical management of LFS represents a real challenge for physicians, considering the wide tumour spectrum associated with germline TP53 mutations, and the diversity of the clinical severity and age of tumour onset observed among families. The 'Toronto protocol' including annual total body MRI and brain MRI, from the first year of life, has recently been elaborated to ensure early tumour detection in germline TP53 mutation carriers (45). In the context of the increased detection of germline TP53 mutations in atypical LFS presentations, our study strengthens the idea that it should be appropriate to adapt the medical management of TP53 mutation carriers to the severity of the mutation.

Materials and Methods

p53 functional assay

This assay was derived from those previously described (15,16). EBV-Immortalized lymphocytes from wild-type TP53 control individuals and from germline TP53 mutation carriers (Table 1) were cultured in RPMI 1640 medium (GIBCO, Invitrogen), supplemented with 10% heat-inactivated fetal bovine serum (Invitrogen) at 37 °C with 5% CO₂. The TP53 genotype was systematically assessed by Sanger sequencing. For the p53 functional assay, cells were seeded in duplicates in 12-well plates (Corning) at a density of 10⁶ cells/well and treated or not with 0.3 μM of doxorubicin (Sigma-Aldrich) for 8 h. After treatment, cells were harvested by centrifugation and total RNA extraction was performed with Nucleospin RNAII kit (Macherey Nagel) according to the manufacturer's instructions and quantified using a ND-1000 UV-Vis Spectrophotometer (NanoDrop technologies).

Reverse transcription (RT) was performed on 100 ng of total RNA using the Verso cDNA synthesis kit (Thermo scientific). A RT-Quantitative Multiplex PCR of Short Fluorescent fragments (RT-QMPSF) was performed to measure the transcriptional induction of 6 p53 target genes selected from transcriptomic analysis (CEP170B [NM_015005], PODXL [MIM*602632, NM_001018111], RRAD [MIM*179503, NM_004165], GLS2 [MIM*606365, NM_013267], CABYR [MIM*612135, NM_012189], TP53I3 [MIM*605171, NM_004881]) using two control amplicons (SF3A1 [MIM*605595, NM_005877] and TBP [MIM*600075, NM_003194]) to normalize the data. Fluorescent amplicons were separated by size on an ABI Prism 3100 Genetic Analyzer (Applied

Biosystems, Foster City, CA), and resulting fluorescent profiles were analysed using the GeneScan 3.7 software (Applied Biosystems). RT-QMPSF profiles from exposed and unexposed cells were superimposed after adjustment of control amplicons to the same height, the fold-change was determined for each p53 target gene and an arbitrary p53 functionality score corresponding to the average of fold-changes was defined (Supplementary Material, Fig. S1). For each TP53 mutation, a functionality score was determined from two independent assays and compared to the scores obtained in TP53 wild-type cells, using two-sided Student's t-test.

Comparative gene expression profiling

Comparative gene expression profiling of EBV-immortalized lymphocytes treated or not with doxorubicin, for 8 h was performed using Whole Human Genome Oligo 4 × 44K Microarray (G4112F, Agilent), according to the Agilent Two-Color Gene Expression workflow, as previously described (15,16). Briefly, starting from 100 ng of total RNA, cRNA was synthesized and labelled using the low input Quick Amp Labeling Kit (Agilent), with Cy3 (for untreated cells) and Cy5 (for treated cells). Following co-hybridization, microarrays were scanned (DNA microarray scanner G2565CA, Agilent Technologies) with a resolution of 5 μm. For each experiment, comparative gene expression profiling was performed in 2 replicates. GeneSpring GX 14.5 software (Agilent Technologies) was used to select the differentially expressed genes between treated and untreated cells in both replicates (Student t-test, P-value < 0.005).

Quantitative real-time PCR

Cells were treated or not with doxorubicin (0.3 μM) for 8 h. For quantitative real-time PCR (qRT-PCR) analysis, total RNA was extracted from cells using a Nucleospin RNAII kit (Macherey Nagel), and cDNA was obtained using the Verso cDNA synthesis kit (Thermo scientific). qPCR was performed using the SsoFast EvaGreen supermix (Bio-Rad) detection method on a CFX96 real-time PCR detection system (Bio-Rad). Three biological replicates were performed for each experiment. The comparative Ct method was used for quantification of MDM2 [MIM*164785, NM_002392], GCNT3 [MIM*606836, NM_004751], LINC01480 [NR_110724], PLXNB1 [MIM*601053, NM_002673] and SRSF8 [MIM*603269, NM_032102] gene expression compared to that of the GAPDH [MIM*138400, NM_002046] gene, used as control. The fold-changes between treated and untreated conditions were then calculated.

Western blot analysis

For total protein extraction, cells were pelleted and homogenized in 100 μl of RIPA buffer (25mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, and 0.1% SDS) from Pierce Biotechnology supplemented with cocktails of protease inhibitors (Sigma-Aldrich) and phosphatase inhibitors (Pierce Biotechnology). Samples were centrifuged at 11,300 g for 20 min at 4 °C to remove cellular debris and the supernatant was collected. Thirty μg of protein were resolved by 10% SDS-PAGE. Proteins were transferred to nitrocellulose membranes (Hybond-C Extra; Amersham Biosciences). The following primary antibodies were used: mouse monoclonal anti-p53 DO-1 antibody (1/2000; Santa Cruz Biotechnology, Inc.), anti-actin JLA20 antibody (1/10000; Sigma Aldrich). Membranes were

incubated with peroxidase-labelled anti-mouse or anti-rabbit antibodies (1/10000) from Jackson Immunoresearch Laboratories, and signals were detected with chemiluminescence reagents using G:Box (Syngene) and GeneSnap software.

p53-binding assay in yeast

Putative p53-binding sites detected by ChIP-Seq were PCR-amplified and sequences required for homologous recombination in yeast were added to PCR products using a second PCR (primer sequences available upon request). The ADE2-YPH500-TP53 yeast strain, containing the pLS76 plasmid and expressing human p53, and the ADE2-YPH500-Empty yeast strain, containing the pSS16 vector, were transformed, using the lithium acetate procedure, with PCR products and a linearized pRS313-cyc-ADE2 vector, containing the ADE2 [NM_001183547.3] open reading frame downstream of a minimal promoter (cyc) (26,46). Transformed yeast cells were plated on a synthetic minimal medium minus leucine and histidine, to select for cells containing the pLS76 or pSS16 and pRS313 vectors, respectively, plus a limited amount of adenine (10 µg/ml) and incubated for 2 days at 30 °C (Supplementary Material, Fig. S4).

Chromatin Immunoprecipitation

EBV-immortalized lymphocytes from wild-type TP53 subjects or from germline TP53 mutation carriers were seeded in duplicates in 15cm dishes, at a density of 20×10^6 cells/dish and treated with 0.3 µM of doxorubicin for 8h (Sigma-Aldrich). Cells were fixed by the addition of 1% formaldehyde for 10 min (Sigma-Aldrich) and then fixation was stopped by the addition of 0.125 M glycine for 5 min (Millipore). Cells were washed 3 times in PBS and then suspended in SDS lysis buffer supplemented with a protease inhibitor cocktail (EZ-ChIP kit, Millipore). Chromatin from 20×10^6 cells was fragmented by sonication using a S220 ultrasonicator (Covaris) yielding genomic DNA fragments with an average size of 250 bp. Chromatin immunoprecipitation was performed with the EZ-ChIP kit (Millipore), according to the manufacturer's instructions, using 4 µg of mouse monoclonal anti-p53 DO-1 antibody (Santa Cruz Biotechnology, Inc.). DNA was purified on Millipore columns (EZ-ChIP kit, Millipore) and then concentrated on columns from DNA clean and concentrator-10 kit (Zymo Research).

High-throughput Sequencing

Libraries were prepared from 4 ng of immunoprecipitated DNA or input DNA, using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® and the NEBNext® Multiplex Oligos for Illumina® (NEB). Briefly, after end-repair, A-tailing and adaptor ligation, libraries were PCR-amplified using 10 cycles and size-selected. Library concentrations and sizes were checked on a TapeStation 2200 system (Agilent). Libraries were then subjected to a 2x75bp paired-end sequencing on an Illumina NextSeq500 sequencer to yield an average of 50 million paired-end reads per sample.

Bioinformatic Analyses

Paired-end 75bp reads were aligned against the Human Reference Genome (assembly hg19, UCSC), using the BWA MEM algorithm (v0.7.10) (47) and sorted with Picard SortSam (v1.141) (<http://broadinstitute.github.io/picard>; date last accessed March

22, 2017). Detection and annotation of the peaks were performed using the HOMER (v4.8) algorithm (48). Peak false discovery rate (FDR) was set to 0.005 (i.e 5 false positive peaks every 1000 called peaks), analysis was performed with default Poisson P-value and fold change parameters, and retained peaks were filtered against the input and against their local background. Peak sequences were checked for significant p53-binding motif enrichment, using HOMER (P-value < 0.05). Each peak was assigned to the gene with the nearest transcription start site (TSS) during the annotation process (HOMER). Intersections of the peaks detected in the different ChIP-Seq experiments were performed using the Odysseus software that we developed. Venn diagrams were plotted using the BioVenn web application (<http://www.cmbi.ru.nl/cdd/biovenn/>; date last accessed March 22, 2017). DAVID (v6.7) (49) was used for the functional annotation of genes with a detectable p53-binding site. The peaks were inspected using the IGV Genome Browser (v2.3) (50).

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We are grateful to Mario Tosi for critical review of the manuscript.

Conflict of interest statement. None declared.

Funding

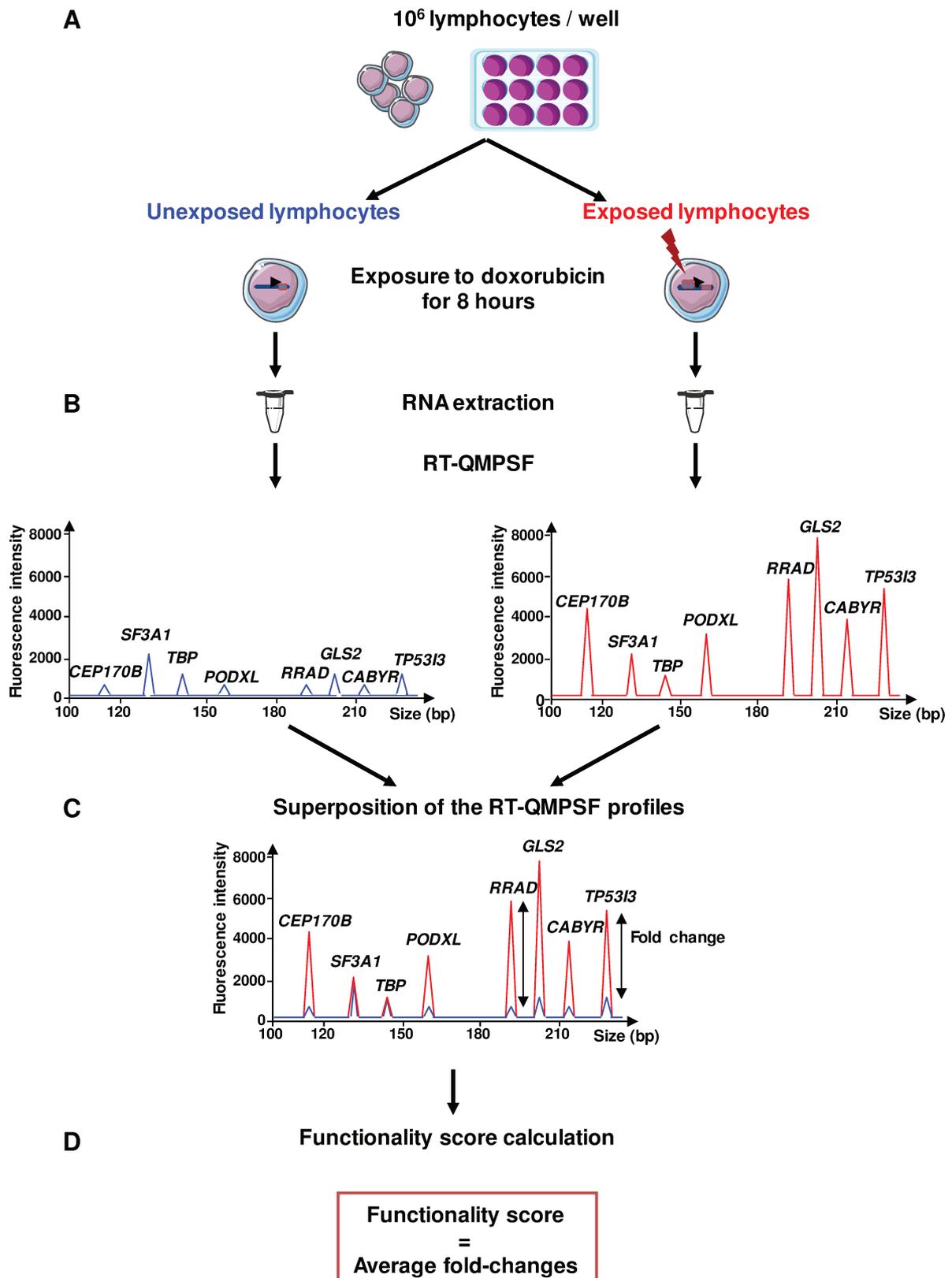
This work was supported by the French National Cancer Institute (INCa), the French Association for Cancer Research (Fondation ARC), the League against cancer (Ligue contre le Cancer) and by the European Regional Development Fund (European Union and Région Normandie). Funding to pay the Open Access publication charges for this article was provided by the French Institute for Health and Medical Research (Inserm).

References

1. Bougeard, G., Renaux-Petel, M., Flaman, J.-M., Charbonnier, C., Fermey, P., Belotti, M., Gauthier-Villars, M., Stoppa-Lyonnet, D., Consolino, E., Brugières, L., et al. (2015) Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *J. Clin. Oncol.*, **33**, 2345–2352.
2. Gonzalez, K.D., Noltner, K.A., Buzin, C.H., Gu, D., Wen-Fong, C.Y., Nguyen, V.Q., Han, J.H., Lowstuter, K., Longmate, J., Sommer, S.S., et al. (2009) Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. *J. Clin. Oncol.*, **27**, 1250–1256.
3. Kamihara, J., Rana, H.Q. and Garber, J.E. (2014) Germline TP53 mutations and the changing landscape of Li-Fraumeni syndrome. *Hum. Mutat.*, **35**, 654–662.
4. Li, F.P., Fraumeni, J.F., Mulvihill, J.J., Blattner, W.A., Dreyfus, M.G., Tucker, M.A. and Miller, R.W. (1988) A cancer family syndrome in twenty-four kindreds. *Cancer Res.*, **48**, 5358–5362.
5. Li, F.P. and Fraumeni, J.F. (1969) Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann. Intern. Med.*, **71**, 747–752.
6. Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischoff, F.Z. and Tainsky, M.A. (1990) Germ line p53 mutations in a familial syndrome

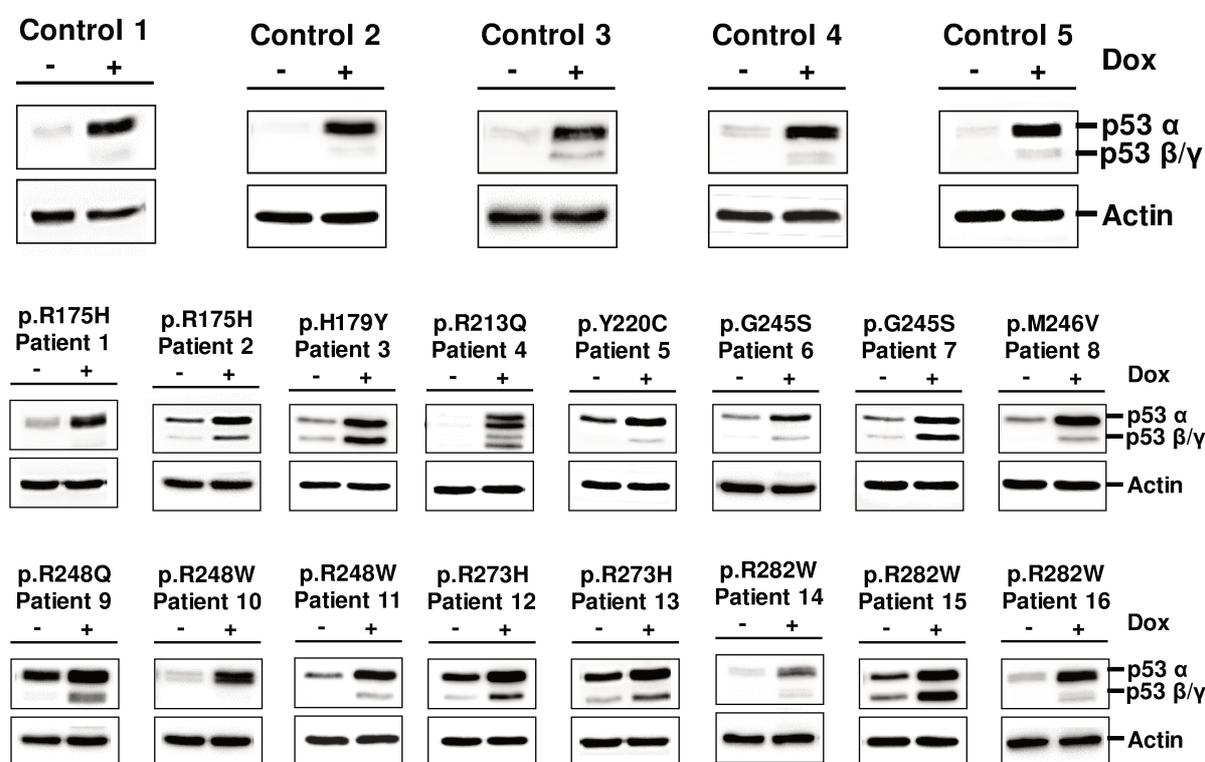
- of breast cancer, sarcomas, and other neoplasms. *Science*, **250**, 1233–1238.
7. Srivastava, S., Zou, Z.Q., Pirollo, K., Blattner, W. and Chang, E.H. (1990) Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature*, **348**, 747–749.
 8. Tinat, J., Bougeard, G., Baert-Desurmont, S., Vasseur, S., Martin, C., Bouvignies, E., Caron, O., Bressac-de Paillerets, B., Berthet, P., Dugast, C., et al. (2009) 2009 version of the Chompret criteria for Li Fraumeni syndrome. *J. Clin. Oncol.*, **27**, 108–109.
 9. Gonzalez, K.D., Buzin, C.H., Noltner, K.A., Gu, D., Li, W., Malkin, D. and Sommer, S.S. (2009) High frequency of de novo mutations in Li-Fraumeni syndrome. *J. Med. Genet.*, **46**, 689–693.
 10. Bougeard, G., Sesboué, R., Baert-Desurmont, S., Vasseur, S., Martin, C., Tinat, J., Brugières, L., Chompret, A., de Paillerets, B.B., Stoppa-Lyonnet, D., et al. (2008) Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *J. Med. Genet.*, **45**, 535–538.
 11. Hanel, W., Marchenko, N., Xu, S., Yu, S.X., Weng, W. and Moll, U. (2013) Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis. *Cell Death Differ.*, **20**, 898–909.
 12. Lang, G.A., Iwakuma, T., Suh, Y.-A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong, L.C., et al. (2004) Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell*, **119**, 861–872.
 13. Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D. and Jacks, T. (2004) Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell*, **119**, 847–860.
 14. Song, H., Hollstein, M. and Xu, Y. (2007) p53 gain-of-function cancer mutants induce genetic instability by inactivating ATM. *Nat. Cell Biol.*, **9**, 573–580.
 15. Zerdoumi, Y., Aury-Landas, J., Bonaïti-Pellié, C., Derambure, C., Sesboué, R., Renaux-Petel, M., Frebourg, T., Bougeard, G. and Flaman, J.-M. (2013) Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. *Hum. Mutat.*, **34**, 453–461.
 16. Zerdoumi, Y., Kasper, E., Soubigou, F., Adriouch, S., Bougeard, G., Frebourg, T. and Flaman, J.-M. (2015) A new genotoxicity assay based on p53 target gene induction. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.*, **789–790**, 28–35.
 17. Petitjean, A., Achatz, M.I.W., Borresen-Dale, A.L., Hainaut, P. and Olivier, M. (2007) TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, **26**, 2157–2165.
 18. Kato, S., Han, S.-Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R. and Ishioka, C. (2003) Understanding the function-structure and function-mutation relationships of p53 tumour suppressor protein by high-resolution missense mutation analysis. *Proc. Natl. Acad.*, **100**, 8424–8429.
 19. Achatz, M.I.W., Olivier, M., Le Calvez, F., Martel-Planche, G., Lopes, A., Rossi, B.M., Ashton-Prolla, P., Giugliani, R., Palmero, E.I., Vargas, F.R., et al. (2007) The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families. *Cancer Lett.*, **245**, 96–102.
 20. Koepfel, M., van Heeringen, S.J., Kramer, D., Smeenk, L., Janssen-Megens, E., Hartmann, M., Stunnenberg, H.G. and Lohrum, M. (2011) Crosstalk between c-Jun and TAp73alpha/beta contributes to the apoptosis-survival balance. *Nucleic Acids Res.*, **39**, 6069–6085.
 21. Riley, T., Sontag, E., Chen, P. and Levine, A. (2008) Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.*, **9**, 402–412.
 22. Zeron-Medina, J., Wang, X., Repapi, E., Campbell, M.R., Su, D., Castro-Giner, F., Davies, B., Peterse, E.F.P., Sacilotto, N., Walker, G.J., et al. (2013) A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell*, **155**, 410–422.
 23. Tebaldi, T., Zaccara, S., Alessandrini, F., Bisio, A., Ciribilli, Y. and Inga, A. (2015) Whole-genome cartography of p53 response elements ranked on transactivation potential. *BMC Genomics*, **16**, 464.
 24. Do, P.M., Varanasi, L., Fan, S., Li, C., Kubacka, I., Newman, V., Chauhan, K., Daniels, S.R., Boccetta, M., Garrett, M.R., et al. (2012) Mutant p53 cooperates with ETS2 to promote etoposide resistance. *Genes Dev.*, **26**, 830–845.
 25. Zhu, J., Sammons, M.A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., Baryste-Lovejoy, D., Al-awar, R., Katona, B.W., Shilatifard, A., et al. (2015) Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature*, **525**, 206–211.
 26. Flaman, J.M., Frebourg, T., Moreau, V., Charbonnier, F., Martin, C., Chappuis, P., Sappino, A.P., Limacher, I.M., Bron, L. and Benhattar, J. (1995) A simple p53 functional assay for screening cell lines, blood, and tumours. *Proc. Natl. Acad. Sci.*, **92**, 3963–3967.
 27. Frebourg, T., Barbier, N., Kassel, J., Ng, Y.S., Romero, P. and Friend, S.H. (1992) A functional screen for germ line p53 mutations based on transcriptional activation. *Cancer Res*, **52**, 6976–6978.
 28. Botcheva, K., McCorkle, S.R., McCombie, W.R., Dunn, J.J. and Anderson, C.W. (2011) Distinct p53 genomic binding patterns in normal and cancer-derived human cells. *Cell Cycle Georget. Tex.*, **10**, 4237–4249.
 29. Botcheva, K. and McCorkle, S.R. (2014) Cell context dependent p53 genome-wide binding patterns and enrichment at repeats. *PLoS One*, **9**, e113492.
 30. Menendez, D., Nguyen, T.-A., Freudenberg, J.M., Mathew, V.J., Anderson, C.W., Jothi, R. and Resnick, M.A. (2013) Diverse stresses dramatically alter genome-wide p53 binding and transactivation landscape in human cancer cells. *Nucleic Acids Res.*, **41**, 7286–7301.
 31. Nikulenkov, F., Spinnler, C., Li, H., Tonelli, C., Shi, Y., Turunen, M., Kivioja, T., Ignatiev, I., Kel, A., Taipale, J., et al. (2012) Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death Differ.*, **19**, 1992–2002.
 32. Smeenk, L., van Heeringen, S.J., Koepfel, M., Gilbert, B., Janssen-Megens, E., Stunnenberg, H.G. and Lohrum, M. (2011) Role of p53 serine 46 in p53 target gene regulation. *PLoS One*, **6**, e17574.
 33. Lieberman, P.M. (2016) Retrotransposon-derived p53 binding sites enhance telomere maintenance and genome protection. *BioEssays*, **38**, 943–949.
 34. Kenzelmann Broz, D., Spano Mello, S., Biegging, K.T., Jiang, D., Dusek, R.L., Brady, C.A., Sidow, A. and Attardi, L.D. (2013) Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. *Genes Dev.*, **27**, 1016–1031.
 35. Tonelli, C., Morelli, M.J., Bianchi, S., Rotta, L., Capra, T., Sabò, A., Campaner, S. and Amati, B. (2015) Genome-wide analysis of p53 transcriptional programs in B cells upon exposure to genotoxic stress in vivo. *Oncotarget.*, **6**, 24611–24626.

36. Ferraro, A. (2016) Altered primary chromatin structures and their implications in cancer development. *Cell. Oncol. Dordr.*, **39**, 195–210.
37. Leroy, B., Girard, L., Hollestelle, A., Minna, J.D., Gazdar, A.F. and Soussi, T. (2014) Analysis of TP53 mutation status in human cancer cell lines: a reassessment. *Hum. Mutat.*, **35**, 756–765.
38. Herrero, A.B., Rojas, E.A., Misiewicz-Krzeminska, I., Krzeminski, P. and Gutiérrez, N.C. (2016) Molecular mechanisms of p53 deregulation in cancer: an overview in multiple myeloma. *Int. J. Mol. Sci.*, **17**, 2003.
39. Varley, J.M., Evans, D.G. and Birch, J.M. (1997) Li-Fraumeni syndrome—a molecular and clinical review. *Br. J. Cancer*, **76**, 1–14.
40. Nicholls, C.D., McLure, K.G., Shields, M.A. and Lee, P.W.K. (2002) Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. Implications for the dominant negative effect. *J. Biol. Chem.*, **277**, 12937–12945.
41. Aramayo, R., Sherman, M.B., Brownless, K., Lurz, R., Okorokov, A.L. and Orlova, E.V. (2011) Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance. *Nucleic Acids Res.*, **39**, 8960–8971.
42. Haupt, S., Raghu, D. and Haupt, Y. (2016) Mutant p53 drives cancer by subverting multiple tumour suppression pathways. *Front. Oncol.*, **6**, 12.
43. Mantovani, F., Walerych, D. and Sal, G.D. (2016) Targeting mutant p53 in cancer: a long road to precision therapy. *Febs J.*, **10.1111/febs.13948**.
44. Sabapathy, K. (2015) The contrived mutant p53 oncogene—beyond loss of functions. *Front. Oncol.*, **5**, 276.
45. Villani, A., Shore, A., Wasserman, J.D., Stephens, D., Kim, R.H., Druker, H., Gallinger, B., Naumer, A., Kohlmann, W., Novokmet, A., et al. (2016) Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. *Lancet Oncol.*, **17**, 1295–1305.
46. Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19–27.
47. Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinform. Oxf. Engl.*, **26**, 589–595.
48. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
49. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
50. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

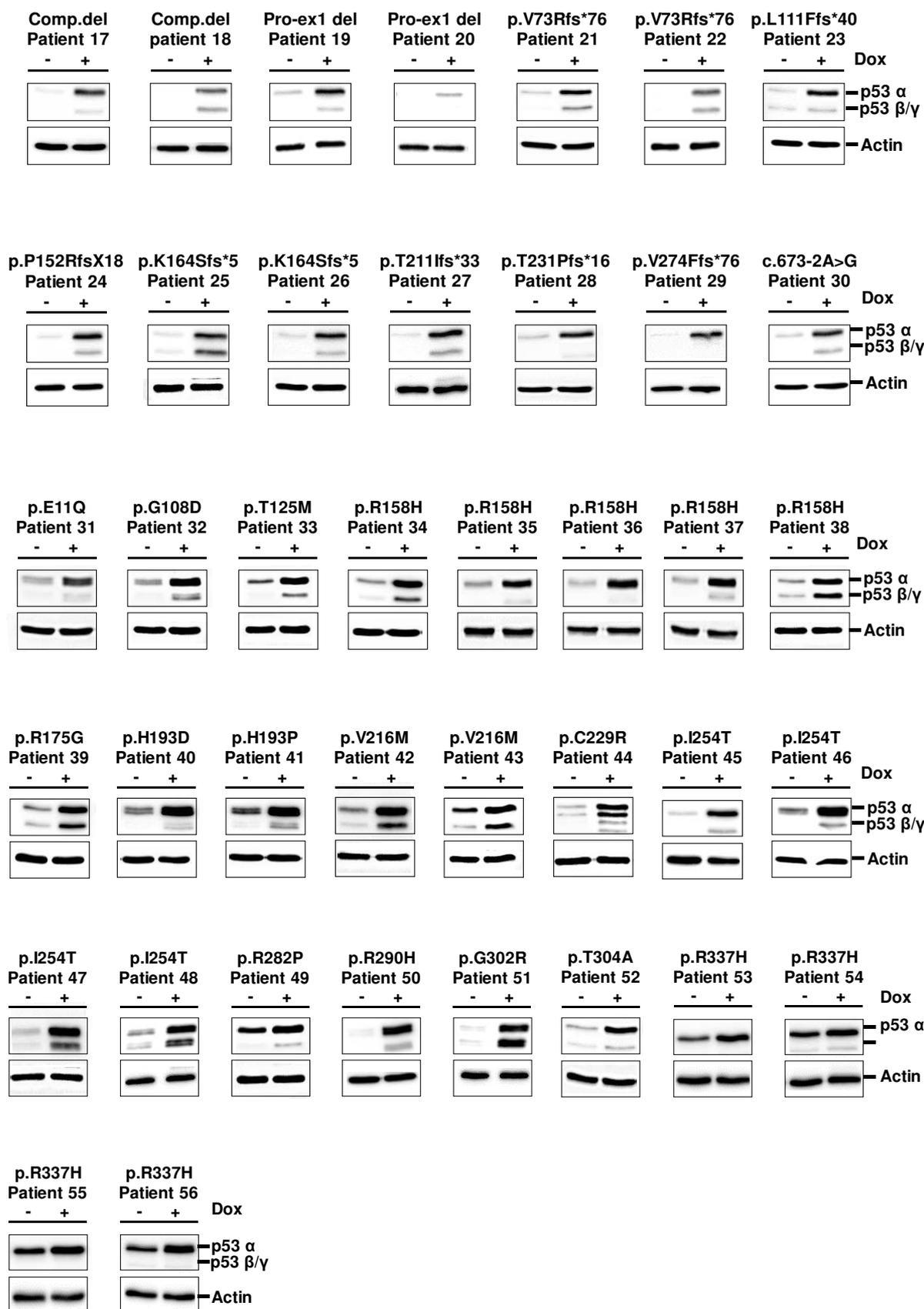


Supplementary Figure S1

Description of the p53 functional assay. (A) Cultured EBV-immortalized lymphocytes from wild-type TP53 control subjects or from germline TP53 mutation carriers are exposed for 8 hours to doxorubicin in order to activate p53. (B) After RNA extraction, expression of selected p53 target genes (CEP170B, PODXL, RRAD, GLS2, CABYR, TP53I3) and control genes (SF3A1 and TBP) is analysed using RT-QMPSF (reverse transcription–quantitative multiplex PCR of short fluorescent fragments). (C) RT-QMPSF profiles generated from exposed and unexposed cells are superimposed after adjustment of the control gene peaks to the same heights in order to determine the fold-change for each p53 target gene. (D) The average of fold-changes is calculated yielding an arbitrary functionality score.

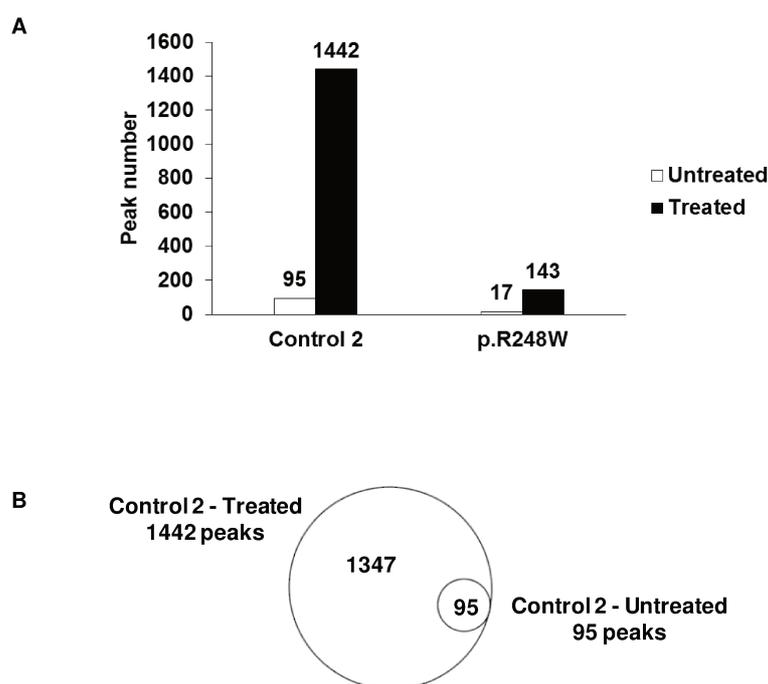


Objectif 1



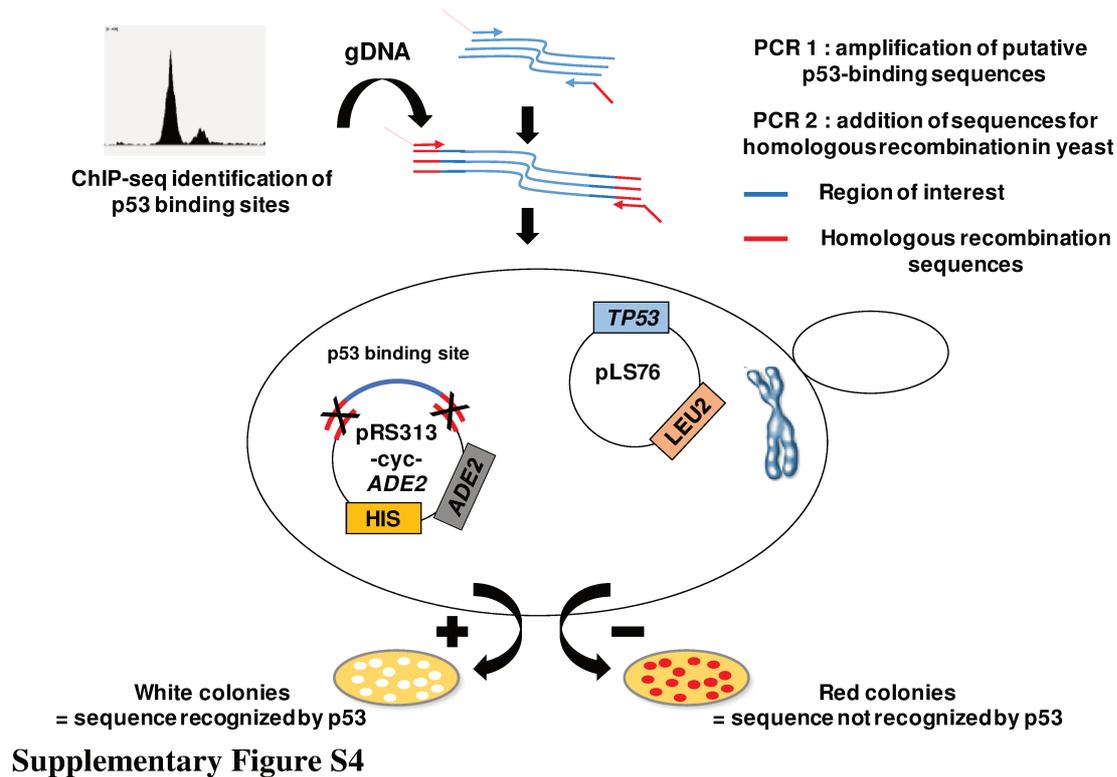
Supplementary Figure S2

p53 stabilisation in EBV-immortalized lymphocytes exposed to doxorubicin. EBV-immortalized lymphocytes from control subjects and germline TP53 mutation carriers were exposed (+) or not (-) to doxorubicin (Dox) during 8 hours. Extracted proteins were separated by SDS-PAGE, and p53 protein was revealed using the DO-1 antibody. The upper band represents the p53 full-length isoform (p53 α) and the lower band the p53 β or γ isoforms. Anti-actin antibody was used as a loading control.

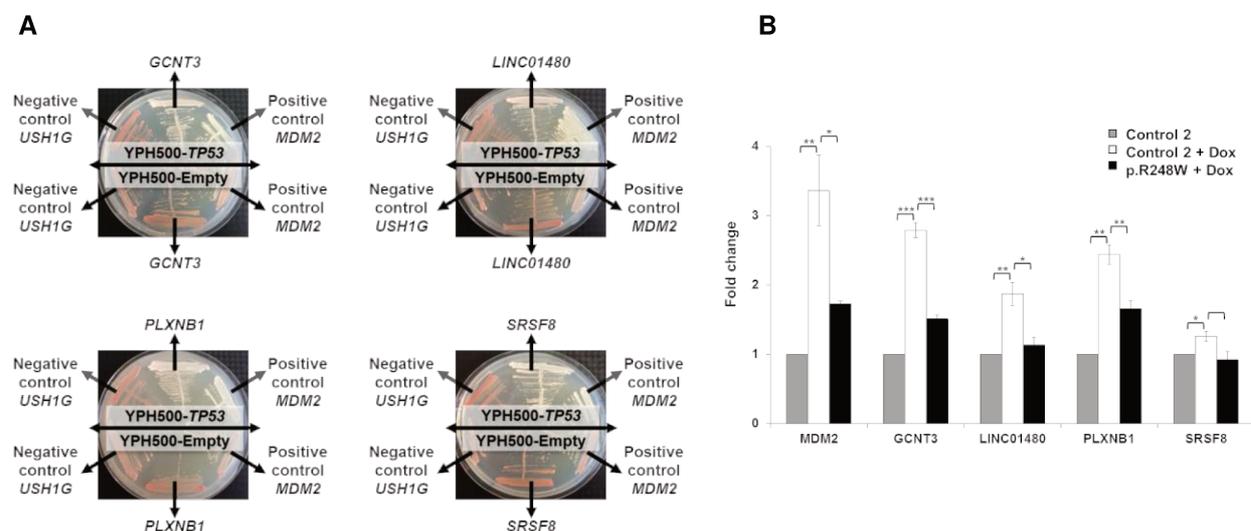


Supplementary Figure S3

Comparison of the p53-binding site number detected by ChIP-Seq between EBV-immortalized lymphocytes exposed and unexposed to doxorubicin. (A) Histograms present the number of peaks detected in common between replicates of wild-type TP53 control lymphocytes (control 2) and lymphocytes harboring the dominant-negative missense TP53 mutation p.R248W. Open bars correspond to untreated cells and black bars to doxorubicin treated cells. (B) Venn diagram illustrating the overlap of peaks detected in wild-type TP53 control lymphocytes (control 2) untreated or treated with doxorubicin for 8h.



p53 binding assay in yeast (see material and methods)



Functional validation, using a functional assay in yeast, of p53-binding sites identified by ChIP-Seq. (A) Results of the p53 binding assay in yeast for 4 new putative p53-target genes (GCNT3, LINC01480, PLXNB1, SRSF8). The MDM2 and USH1G genes were used as positive and negative control, respectively. Results obtained in yeast strain YPH 500-TP53 expressing human p53 and yeast strain YPH 500-Empty containing the empty vector are shown. (B)

Objectif 1

Quantitative real-time PCR validation of newly identified putative p53 targets in EBV-immortalized lymphocytes. mRNA levels were quantified in cells untreated and treated with doxorubicin (Dox) for 8h and normalized to GAPDH. Fold change of gene expression in response to treatment are shown for the control lymphocytes (control 2) and for the p.R248W mutant lymphocytes. Expression of the MDM2 gene was used as a positive control. Values are the average of three independent experiments; the error bars represent mean \pm SEM, and the significance was determined by unpaired t-test (* $p < 0.05$ and ** $p < 0.01$; *** $p < 0.001$).

III- Discussion

Grâce à notre collection exceptionnelle de lymphocytes immortalisés dérivés de patients LFS et à la simplification de l'essai fonctionnel développé, nous avons montré que les mutations constitutionnelles faux-sens de TP53, associées à un effet dominant négatif, altèrent considérablement la réponse de p53 aux lésions de l'ADN à l'état hétérozygote en comparaison avec les mutations nulles associées à une perte de fonction exclusive. De plus, notre analyse de CHIP-Seq nous a permis de montrer un défaut majeur de la fixation de p53 à l'ADN dans les lymphocytes du patient LFS porteur d'une mutation faux-sens associée à un effet dominant négatif. Ce défaut de fixation à l'ADN de la protéine p53 mutante explique probablement l'altération drastique de la réponse transcriptionnelle de p53 aux lésions de l'ADN dans les lymphocytes de patients LFS porteurs d'une mutation faux-sens avec un effet dominant négatif.

Ainsi, nos résultats confirment que la sévérité clinique et la prédominance des mutations faux-sens avec un effet dominant négatif dans le syndrome de Li-Fraumeni résulte d'une extrême sévérité biologique due à un défaut majeur de fixation de p53 à l'ADN en réponse aux lésions de l'ADN, ce qui a pour conséquence la perte de l'activité transcriptionnelle de p53 qui est essentielle à son activité onco-suppressive. Cet effet drastique que nous avons observé avec des mutations faux-sens explique probablement la prédominance de ce type de mutations observée au niveau somatique dans les tumeurs et au niveau constitutionnel dans le LFS. Cet essai réalisé à partir de lymphocytes, représente donc un nouvel outil puissant d'une part pour appréhender la corrélation génotype - phénotype dans le LFS, d'autre part pour classer des variations de signification biologique inconnue et enfin pour détecter des patients présentant une inactivation de la voie p53, sans mutation identifiée.

Objectif 2

Recherche de nouveaux biomarqueurs indicateurs de la présence d'une altération constitutionnelle des gènes suppresseurs de tumeurs *TP53* et *BRCA* par approche multi-omique

I- Contexte scientifique

L'identification de la mutation constitutionnelle à l'origine d'une prédisposition génétique au cancer est essentielle à la prise en charge médicale des apparentés et de leur famille. L'implémentation récente des techniques de séquençage à haut-débit à la routine diagnostique de ces pathologies a permis une analyse extensive des gènes connus pour leur implication dans ces syndromes et a également permis de rechercher de nouvelles bases moléculaires en séquençant d'autres gènes candidats en parallèle. Aujourd'hui, le défi ne réside plus dans la détection de ces altérations génétiques, mais dans leur interprétation biologique et clinique et leur classification. La question de l'interprétation de la variation est particulièrement cruciale lorsqu'elle conditionne la stratégie thérapeutique. Il est donc essentiel de disposer de tests simples, adaptables en routine diagnostique pour faciliter l'interprétation des variations génétiques et les classer en variants pathogènes ou neutres. Notre unité de recherche s'est spécialisée depuis des années sur l'interprétation fonctionnelle de ces VSI notamment par le développement de tests fonctionnels pour tester l'impact de ces VSI sur l'épissage (Tournier *et al.*, 2008; Gaildrat *et al.*, 2010; Soukarieh *et al.*, 2016), mais également sur des tests fonctionnels simples comme ceux que nous avons récemment développés pour évaluer l'impact de mutations constitutionnelles du gène *TP53* sur l'activité transcriptionnelle de la protéine p53 (Zerdoumi *et al.*, 2013, 2017) (Résultat objectif 1).

En parallèle du séquençage à haut-débit de génomes, d'autres approches omiques se sont développées pour l'analyse à haut-débit du transcriptome, du protéome, du métabolome ou encore de l'épigénome. Le développement des technologies omiques offre aujourd'hui une opportunité sans précédent pour identifier de nouveaux biomarqueurs spécifiques d'une condition physiologique donnée. Nous avons choisi pour la deuxième partie de mon projet de thèse d'évaluer la pertinence de ces approches omiques seules ou en combinaison pour identifier de nouveaux biomarqueurs capables de discriminer une cellule sans mutation d'une cellule avec mutation délétère.

Nous avons choisi comme première approche une **analyse métabolomique non ciblée**. Conceptuellement, la métabolomique trouve son origine dans les pratiques des anciens médecins grecs qui utilisaient les caractéristiques organoleptiques de l'urine pour le diagnostic. Par exemple, la douceur urinaire révèle les niveaux élevés de glucose dans le diabète. Ces caractéristiques chimiques organoleptiques sont bien entendu liées au métabolisme. Olivier et collaborateurs en 1998 ont défini le métabolome comme l'ensemble

des métabolites synthétisés par un organisme (Oliver *et al.*, 1998). Le métabolome désigne ainsi tous les métabolites présents dans un système biologique, un fluide, une cellule ou un tissu donné à un instant T (Nicholson *et al.*, 1999). Ces métabolites sont de petites molécules produites par des réactions enzymatiques. L'analyse métabolomique non ciblée est basée sur les caractérisations biochimiques et moléculaires du métabolome et sur les modifications des métabolites liés aux variables génétiques, environnementales, médicamenteuses ou alimentaires, en plus d'autres facteurs (Fiehn, 2002; Holmes *et al.*, 2008; Dunn *et al.*, 2011; Benton *et al.*, 2012). Cette approche prometteuse est encore très peu explorée en génétique des cancers. Le métabolome représente cependant un réservoir important de biomarqueurs potentiels. De plus, avec le développement d'approches de métabolomique en temps-réel, réalisées au sein même du bloc opératoire, le profil métabolomique des individus avec prédisposition au cancer nous semble très intéressant à analyser. En effet le couteau intelligent (iKnife) représente une avancée significative dans l'échantillonnage *in vivo* et la métabolomique en temps réel. Ce processus vaporise les tissus et la fumée qui en résulte est transférée dans un spectromètre de masse afin de fournir l'analyse en temps réel qui facilite la prise de décision clinique dans la salle d'opération (Balog *et al.*, 2013).

Plusieurs techniques d'analyse et stratégies de préparation des échantillons sont nécessaires pour récupérer la majeure partie du métabolome. Nous avons choisi d'appliquer une approche métabolomique non ciblée qui repose principalement sur l'analyse qualitative ou semi-quantitative du plus grand nombre possible de métabolites contenus dans un échantillon biologique pour répondre aux questions biologiques. Plusieurs plates-formes analytiques permettent l'évaluation semi-quantitative (intensités relatives) des métabolites. Nous avons choisi d'analyser nos échantillons par spectrométrie de masse à haute résolution couplée à la mobilité ionique et à la chromatographie liquide ultra-haute performance (UHPLC-IM-MS). Ces différentes méthodes de séparation améliorent la sensibilité, la couverture, la détection des métabolites et fournissent des informations moléculaires complémentaires et orthogonales. Ainsi, cette stratégie analytique offre une combinaison multidimensionnelle puissante pour l'identification des métabolites et, par conséquent, un outil prometteur pour la découverte de biomarqueurs, ébauche préliminaire de futurs tests prédictifs.

Comme deuxième approche, nous avons choisi pour ce projet de réaliser des analyses par séquençage à haut-débit du transcriptome complet ou **RNA-Seq total**. En effet, le développement du séquençage de nouvelle génération à haut-débit (NGS) a révolutionné la transcriptomique en permettant l'analyse de l'ARN via le séquençage de l'ADN

complémentaire (ADNc) (Wang *et al.*, 2009). Cette méthode, appelée séquençage ARN (RNA-Seq) a révolutionné notre compréhension de la nature complexe et dynamique du transcriptome. Le RNA-Seq fournit une vue plus détaillée et quantitative de l'expression génique qui est un processus largement étudié et un axe majeur de la génomique fonctionnelle. L'expression génétique concerne le flux d'informations génétiques de la matrice d'ADN génomique vers les produits protéiques fonctionnels. Ainsi, cet outil est rapidement devenu un test d'expression génique standard, en particulier pour interroger l'abondance et la diversité des transcrits.

Le RNA-Seq a permis d'importantes découvertes dans de nombreux domaines de recherche et spécifiquement dans le cancer et fait l'objet de diverses applications, comme par exemple l'analyse différentielle de l'expression génique que nous avons choisi d'appliquer, et peut fournir de nouvelles informations sur l'épissage alternatif, les exons non annotés et les nouveaux transcrits. Il a aussi permis l'identification de dizaines de milliers de nouveaux longs ANR non codants ou lincRNA dans différents tissus (Trapnell *et al.*, 2010; Cabili *et al.*, 2011; Ching *et al.*, 2014). Ainsi, le projet Encyclopedia of DNA Elements (ENCODE) a révélé qu'environ 62 % de la totalité du génome est transcrite en longues séquences d'ARN (> 200 paires de bases) (ENCODE Project Consortium, 2012). Étant donné que 3 % du génome code pour les exons codant pour les protéines, la grande majorité de ces transcrits sont des ARN non codants (lncRNAs). Parmi ces lncARN, environ un tiers provient des régions intergéniques (lincRNAs) (ENCODE Project Consortium, 2012). Pour des raisons de facilité d'interprétation, les analyses de transcriptomes courantes ne concernent que les transcrits polyadénylés (mRNA) et sont généralement centrées sur les transcrits codant pour une protéine. Mais compte tenu du rôle grandissant des ARN non codants dans la réponse aux lésions de l'ADN (Khanduja *et al.*, 2016), nous avons choisi de les inclure dans notre analyse pour pouvoir détecter des variations d'expression de transcrits pouvant correspondre à des ARN non codants non polyadénylés. L'objectif étant d'identifier des biomarqueurs différentiellement exprimés entre la condition sauvage et mutée, le manque de connaissance sur la fonction de ces transcrits encore mal connus ne représente pas un obstacle à la réalisation de ce projet. L'utilisation du RNA-Seq pour étudier le transcriptome permet non seulement de détecter les variations quantitatives des transcrits, mais également les variations qualitatives résultant d'épissages alternatifs. Les transcrits différentiellement exprimés entre les cellules sauvages et les cellules mutées à l'état hétérozygote pour *TP53* ou *BRCA* seront

sélectionnés comme biomarqueurs. Les données issues de ces deux approches omiques pourront dans un second temps être intégrées.

Nous avons appliqué cette approche à deux formes mendéliennes de cancer : - le syndrome de Li-Fraumeni et la prédisposition aux cancers du sein et de l'ovaire. Pour l'une et l'autre de ces pathologies, l'apparition de tumeurs chez les patients nécessite le plus souvent l'inactivation au niveau somatique du deuxième allèle de ces gènes impliqués dans la réponse aux lésions de l'ADN. Les cellules sont alors dans l'incapacité de réparer les cassures double brin de l'ADN dans le cas des gènes BRCA et dans l'incapacité d'activer une réponse transcriptionnelle appropriée aux lésions de l'ADN dans le cas du gène TP53. Cette fragilité va permettre l'accumulation de nouvelles altérations au sein des cellules mutées aboutissant à l'activation somatique d'oncogènes ou à l'inactivation de gènes suppresseurs de tumeurs. Si l'apparition de tumeurs chez ces patients nécessite un deuxième évènement somatique, la présence d'une mutation constitutionnelle à l'état hétérozygote chez un patient est probablement suffisante pour entraîner un phénotype détectable à l'aide d'outils moléculaires spécifiques. Notre stratégie afin de démasquer cet endophénotype réside dans le fait d'appliquer des drogues génotoxiques, pour ainsi exacerber les différences existantes entre les lymphocytes de sujets témoins et les lymphocytes des patients porteurs d'une mutation à l'état hétérozygote et mettre en évidence une réponse différentielle aux lésions de l'ADN entre ces génotypes.

Notre premier objectif était de démontrer sur la base de ces approches omiques seules ou en combinaison, la possibilité de différencier les groupes de mutants du groupe des témoins. Comme deuxième objectif, sur la base de ces données omiques obtenues, nous avons souhaité développer un test simple, rapide et adaptable à la routine de diagnostic, en intégrant les biomarqueurs discriminants et différentiellement exprimés entre les deux groupes dans un test de RT-MLPA (*Reverse Transcription Multiplex Ligation-dependant Probe Amplification*) ou RT-QMPSF (*Reverse Transcription Quantitative PCR of Short Fluorescent Fragments*).

II- Matériels et Méthodes

1- Culture cellulaire et traitements génotoxiques

❖ Lignées cellulaires

Les lignées utilisées ont été isolées à partir de lymphocytes de patients et immortalisées par infection au virus Epstein-Barr (EBV). Brièvement, les cellules mononucléées du sang périphérique PBMC (Peripheral Blood Mononuclear Cell) ont été isolées du sang total par séparation sur gradient de Ficoll, puis les cellules B ont été infectées par le virus d'Epstein-Barr (EBV) produit par une lignée de cellules de marmouset. Après 8 semaines de culture, les lymphocytes immortalisés par EBV ont été congelés dans l'azote liquide pour une conservation à long terme.

Pour cette étude les lignées suivantes ont été sélectionnées :

- Quatre lignées lymphoblastoïdes établies à partir de patients Li-Fraumeni :
3 porteurs d'une mutation faux-sens à l'état hétérozygote du gène TP53 avec effet dominant négatif (NM_000546.4 : p.(Arg175His), p.(Arg248Trp), p.(Arg273His) et une mutation nulle (NM_000546.4 : c.(?_-202)(*1207_?)del ; délétion totale).
- Quatre lignées issues de patientes avec mutation constitutionnelle hétérozygote de BRCA1 (NM_007294.3 : c.81-80_547+81dup ; NM_007294.3 : c.1-?_80+?del) ou BRCA2 (NM_000059.3: c.5993_5994del ; NM_000059.3: c.7008-1G>A)
- Ces lignées seront comparées à quatre lignées EBV témoins (T401 ; T403 ; T404 ; T-SK) établies à partir de donneurs indemnes de cancer (**Tableau 4**).

❖ Génotypage des lignées par séquençage à haut-débit d'un panel de gènes

La présence de la mutation délétère des gènes TP53, BRCA1 ou BRCA2 a été vérifiée par séquençage à haut-débit d'un panel de 201 gènes impliqués dans le cancer. La présence d'autres mutations clairement délétères a été écartée. Les quatre lignées témoins ont également été séquencées pour vérifier l'absence de mutations dans les gènes étudiés. Après l'extraction de l'ADN selon les instructions du kit QuickGene DNA whole blood Kit (DB-L) (Kurabo, Osaka, Japan) le dosage a été réalisé par fluorimétrie sur le Berthold Twinkle LB970 (Grosseron, Coueron, France). Les librairies d'ADN ont été effectuées manuellement ou de manière robotisée selon le protocole SureSelect^{QXT} d'Agilent. Brièvement, 50 ng d'ADN ont

Objectif 2

Identifiant	Type Lignée	Génotype	Nomenclature ^d	Exon ^e	Type de Mutation
T401	Témoin	WT / WT	WT	/	/
T403	Témoin	WT / WT	WT	/	/
T404	Témoin	WT / WT	WT	/	/
T-SK	Témoin	WT / WT	WT	/	/
R248W	<i>TP53</i> muté ^a	MT / WT	c. 742C>T ; p.(Arg248Trp)	7	Mutation faux sens avec effet dominant négatif ^f
R273H	<i>TP53</i> muté ^a	MT / WT	c.818G>A; p.(Arg273His)	8	Mutation faux sens avec effet dominant négatif ^f
R175H	<i>TP53</i> muté ^a	MT / WT	c. 524G>A ; p.(Arg175His)	5	Mutation faux sens avec effet dominant négatif
Del-totale	<i>TP53</i> muté ^a	MT / WT	c.(?_-202)_(*1207_?)del; p.0	/	Mutation perte de fonction
1823	<i>BRCA1</i> muté ^b	MT / WT	c.1-?_80+?del ; p.?	2	Mutation perte de fonction
827	<i>BRCA1</i> muté ^b	MT / WT	c.81-80_547+81dup; p.?	3-7	Mutation perte de fonction
1820	<i>BRCA2</i> muté ^c	MT / WT	c.5993_5994del ; p.(Gln1998Argfs*4)	11	Mutation perte de fonction
1838	<i>BRCA2</i> muté ^c	MT / WT	c.7008-1G>A ; p.?	14	Mutation perte de fonction

Tableau 4 : Lignées de lymphocytes immortalisés des patients utilisées pour les différentes approches expérimentales

a : NM_000546.4 ; b : NM_007294.3 ; c : NM_000059.3

d : Nomenclature des mutations au niveau de l'ADN complémentaire et conséquences prédites au niveau protéique ; e : Numéro de l'exon ou de l'intron où se situe la mutation ; f : Selon la base de données de l'IARC et d'après Zerdoumi et al., 2017 ; WT : Wild Type ; MT : Mutant.

été fragmentés et indexés par tagmentation (utilisation de transposase) permettant leur séquençage en pool. Les fragments d'intérêt ont été capturés en phase liquide à l'aide des sondes Agilent, puis séquencés par séquençage paired-end 2x150 pb sur séquenceur NextSeq 500 (Illumina, Eindhoven, The Netherlands) de notre unité.

Une analyse bioinformatique a été réalisée pour détecter les variations ponctuelles, les délétions/insertions, les CNV. Le pipeline bioinformatique employé est basé sur l'utilisation des outils BWA et GATK pour effectuer les étapes d'alignement et de variant calling comme recommandé par le Broad Institute (best practice). L'étape de variant calling est réalisée par un double pipeline : GATK HaplotypeCaller afin de détecter les mutations à l'état hétérozygote ou homozygote et VarScan2 pour détecter les mutations en mosaïques. L'annotation des variants a été réalisée par Alamut Batch et par Anovar. Les variations du nombre de copies (CNV) ont été également interrogées à l'aide de l'outil CANOES.

❖ Culture et traitement des lignées

Après décongélation, les lignées ont été maintenues dans du milieu RPMI 1640 (GIBCO, Life Technologies, Carlsbad, CA), additionné de 10 % de sérum de veau fœtal (Invitrogen, Life Technologies, Carlsbad, CA) et 1 % de L-glutamine (Invitrogen, Life Technologies, Carlsbad, CA) à 37°C avec 5 % de CO₂. Dans un premier temps, les cellules ont étéensemencées en double pour chaque approche dans des plaques 12 puits (Corning, New York, United States) à une densité de 10⁶ cellules / puits. Ensuite, les cellules ont été traitées ou non avec 200 ng/ml (0,3 µM concentration finale) de doxorubicine (Sigma Aldrich, St. Louis, Etats-Unis) pour les mutants TP53 et 10,5 µg/mL (7 µM concentration finale) de bléomycine (Enzo, Life Sciences, UK) pour les mutants BRCA pendant 8 heures avant récupération des cellules.

2- Analyse métabolomique

❖ Extraction des métabolites

Une extraction des métabolites a été réalisée à partir de nos lignées traitées ou non traitées. Dans l'objectif d'une analyse globale et non ciblée, l'extraction a été réalisée pour isoler le maximum de métabolites polaires possible. Dans un premier temps, les cellules ont été récupérées dans un tube, et centrifugées 10 minutes à 1500 tours/mn. Le surnageant a été éliminé par retournement rapide et 2 lavages ont été réalisés avec 2 ml de NaCl à 0.9 %. Les culots ont ensuite été séchés 30 minutes à 1 heure.

Les cellules ont ensuite été immédiatement congelées à - 80°C pour éviter une dégradation possible des métabolites. Les échantillons ont été par la suite traités avec 500 µL de méthanol froid (VWR Chemicals, France) par culot cellulaire (500 µL/tube), vortexés soniqués pendant 2 secondes. Puis les échantillons ont été laissés sur la glace pendant 15 minutes pour permettre la précipitation des protéines et centrifugés pendant 10 minutes à 14 000 g à 4°C. 300 µL du surnageant de chaque échantillon ont été transférés et séchés sous courant d'azote. Les extraits séchés ont été mis en suspension avec 100 µL de phase mobile (acétonitrile/eau 50/50) avant analyse. Ensuite, en poolant 10 µL de chaque échantillon suspendu un échantillon de contrôle a été préparé en diluant ce pool au 1/2 et au 1/4.

❖ **Analyse par spectrométrie de masse couplée à la chromatographie liquide haute performance et à la mobilité ionique (UHPLC-IM-MS)**

Les échantillons ont été analysés par UHPLC-IM-MS sur un système Waters nanoACQUITY UPLC (Waters, Milford, USA) couplé à un spectromètre de masse à un temps de vol hybride avec un quadripôle et couplé à une cellule de mobilité ionique (SYNAPT G2 HDMS, Waters MS Technologies, Manchester, Royaume-Uni) et le volume d'injection pour chaque échantillon était 2 µL. Les métabolites ont été séparés sur une colonne de 75 µm x 100 mm x 1,7 µm HSS C18 et precolonne de piège de 180 µm x 20 mm x 5 µm HSS C18. Une analyse technique en triple a été effectuée de manière aléatoire, avec des échantillons de contrôle de qualité analysés toutes les cinq injections.

❖ **Analyse bioinformatique, statistique et contrôle qualité**

A- Traitement et analyse des données de métabolomique

Une étape de filtration des données a été effectuée afin d'éliminer le bruit de fond acquis avec les données, puis une étape de détection des pics a été effectuée. Cette étape consiste à convertir les données brutes en données discrètes sous forme de pic de sorte que chaque ion soit représenté par un pic. Après la détection des pics, un alignement sur tous les profils a été nécessaire pour générer un ensemble combiné de caractéristiques interprétables. Les données générées ont été traitées par MassLynx et DriftScope (Waters, Milford, USA).

Après l'évaluation de la qualité des données, la normalisation des données afin de supprimer les biais expérimentaux a été réalisée par l'approche SVR (Support Vector Regression) suivie de la transformation de ces données avec les deux méthodes Log-

Transformation et Pareto-Scaling afin de modifier la distribution des données et les préparer aux analyses statistiques ultérieures.

B- Analyse statistique multivariée des données métabolomiques

Deux analyses statistiques multivariées ont été effectuées sur toutes les données à l'aide du logiciel de statistiques SIMCA (Umetrics, Sweden) :

- Une analyse en composantes principales (**ACP**) qui est une analyse multivariée descriptive **non-supervisée** permettant une visualisation en 2D ou 3D des données pour une analyse plus intuitive du regroupement des différents échantillons.
- Une analyse **OPLS-DA** (Orthogonal Partial Least Square Discriminant Analysis) qui est une analyse multivariée explicative **supervisée**. L'objectif de cette méthode est d'accentuer la séparation entre les deux groupes définis et de déterminer les variables qui portent l'information permettant la séparation des deux classes. La validation des modèles se fera par validation interne croisée et validation externe. Les métabolites les plus discriminants (biomarqueurs potentiels) seront annotés.

3- Analyse transcriptomique par RNA-Seq global

❖ Extraction des ARN

Après le traitement génotoxique, les cellules ont été récoltées par centrifugation et l'extraction de l'ARN total a été réalisée avec le kit Nucleospin RNAII (Macherey Nagel, Düren, Germany) selon les instructions. Les ARN ont été quantifiés à l'aide d'un UV-VIS ND-1000 spectrophotomètre (technologies de Nanodrop) du fabricant. Tous les échantillons d'ARN utilisés dans cette étude avaient un ratio de 260/280 supérieur à 1,9. La quantité d'ARN a été estimée en utilisant un kit Qubit RNA BR Assay (Invitrogen) et l'intégrité de l'ARN a été évaluée sur une TapeStation2200 (Agilent, Santa Clara, United States) qui produit un indice d'intégrité de l'ARN (RIN) compris entre 1 et 10, sachant que l'indice 10 représente les échantillons de meilleure qualité présentant la moindre dégradation. Le RIN estime l'intégrité de l'échantillon en utilisant l'électrophorèse sur gel et l'analyse des ratios des bandes ribosomiques de 28S à 18S. Seuls les échantillons avec RIN supérieur à 8,0 ont été séquencés. L'ARN a été congelé et stocké à -80°C immédiatement après la production.

❖ Vérification de l'induction d'un stress génotoxique par RT-QMPSF

La transcription inverse (RT) a été réalisée sur 100 ng d'ARN total en utilisant le kit de synthèse d'ADN complémentaire Verso (Thermo scientific, Waltham, United States). Puis une PCR multiplexe quantitative de courts fragments fluorescents (RT-QMPSF) a été réalisée pour mesurer l'induction de la transcription des 6 gènes cibles de p53 sélectionnés dans l'essai fonctionnel développé (*KIAA0284* [NM_015005], *PODXL* [MIM * 602632, NM_001018111], *RRAD* [MIM * 179503, NM_004165], *GLS2* [MIM * 606365, NM_013267], *CABYR* [MIM * 612135, NM_012189], *TP53I3* [MIM * 605171, NM_004881]) et en utilisant deux amplicons témoins (*SF3A1* [MIM * 605595, NM_005877] et *TBP* [MIM * 600075, NM_003194]) pour normaliser les données. Les amplicons fluorescents ont été séparés par taille sur un analyseur génétique ABI Prism 3500 (Applied Biosystems, Foster City, CA) et les profils fluorescents résultants ont été analysés à l'aide du logiciel GeneScan 3.7 (Applied Biosystems). Les profils de RT-QMPSF à partir de cellules exposées et non exposées ont été superposées après ajustement des amplicons contrôles à la même hauteur, le facteur de variation est déterminé pour chaque gène cible de p53 et un score arbitraire de la fonctionnalité de p53 correspondant à la moyenne des facteurs de variation a été défini.

Nous avons effectué une analyse de RNA-Seq total en triplicats biologiques. Nous avons tout d'abord procédé à une déplétion des ARN ribosomaux à l'aide du kit *NEBNext rRNA Depletion* (New England Biolabs, Massachusetts, États-Unis) en partant d'1 µg d'ARN total (**Figure 34**). En effet, l'ARN ribosomal (ARNr) est le composant le plus abondant de l'ARN, comprenant la majorité (> 80 % à 90 %) des molécules présentes dans un échantillon d'ARN total. L'élimination de l'ARN ribosomal a été vérifiée par une analyse sur le système Agilent 2200 TapeStation. Ensuite, la construction des bibliothèques a été effectuée à l'aide du kit *NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB)* qui repose sur une fragmentation chimique de l'ARN, une étape de production de cDNA à l'aide d'amorces aléatoires et l'ajout d'adaptateurs permettant l'amplification clonale et le séquençage (**Figure 35**). Pour l'étape de séquençage, 8 échantillons ont été poolés par run. Le séquençage a été réalisé à l'aide d'un séquençage Illumina Nextseq-500 pour générer 50M de read pairés en moyenne par échantillons.

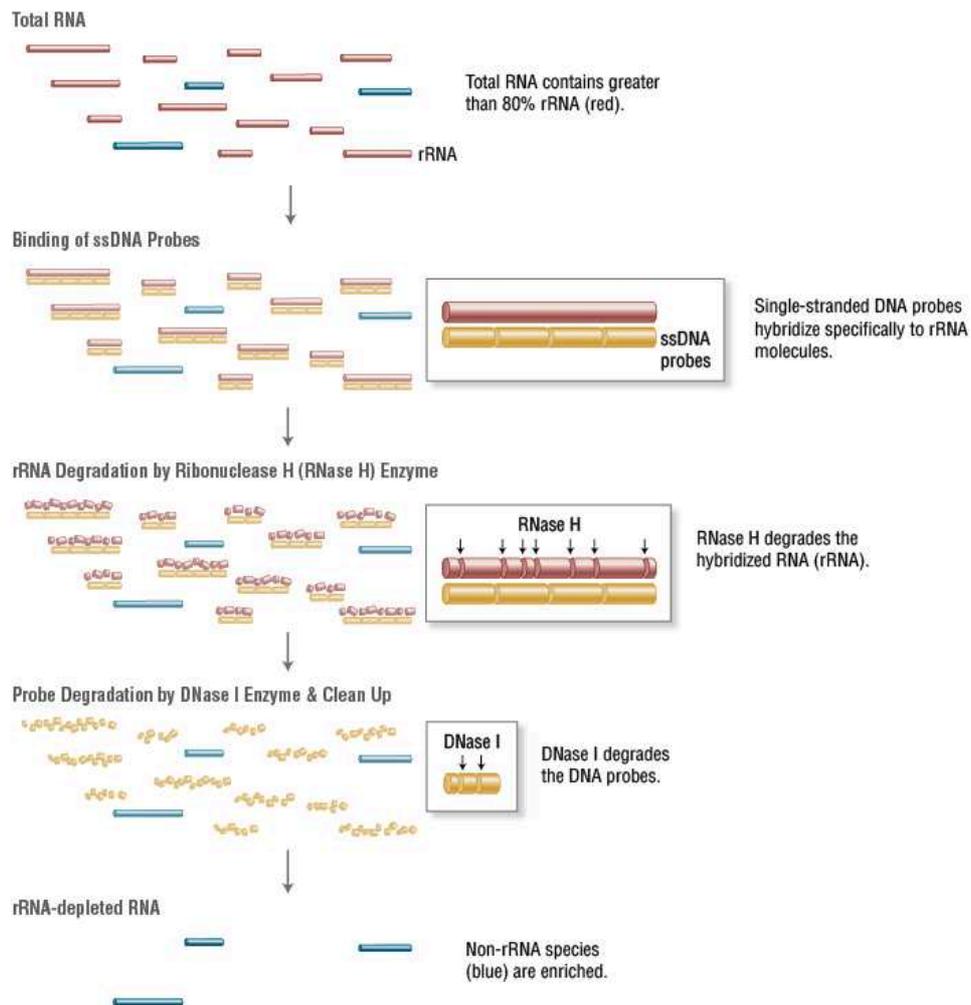


Figure 34 : Déplétion des ARN ribosomiques

Le kit de déplétion d'ARNr NEBNext (humain / souris / rat) utilise une méthode basée sur la RNaseH pour épuiser à la fois l'ARN cytoplasmique (ARNr 5S, ARNr 5.8S, ARNr 18S et ARNr 28S) et l'ARN ribosomal mitochondrial (ARNr 12S et ARNr 16S) provenant de l'ARN total humain. Des sondes d'ADN anti-sens courtes (50 à 80 bases) sont construites et sont complémentaires aux séquences des ARN cibles. Ces séquences forment des hybrides ARN : ADN sur toute la longueur des espèces d'ARN. Ensuite, un traitement par RNaseH suivi de DNaseI détruit efficacement les espèces d'ARN ciblées ainsi que les sondes d'ADN résiduelles. Ce processus peut être facilement reconfiguré pour cibler différentes espèces d'ARN, préserve l'intégrité des ARN non ciblés et peut être effectué en environ deux heures.

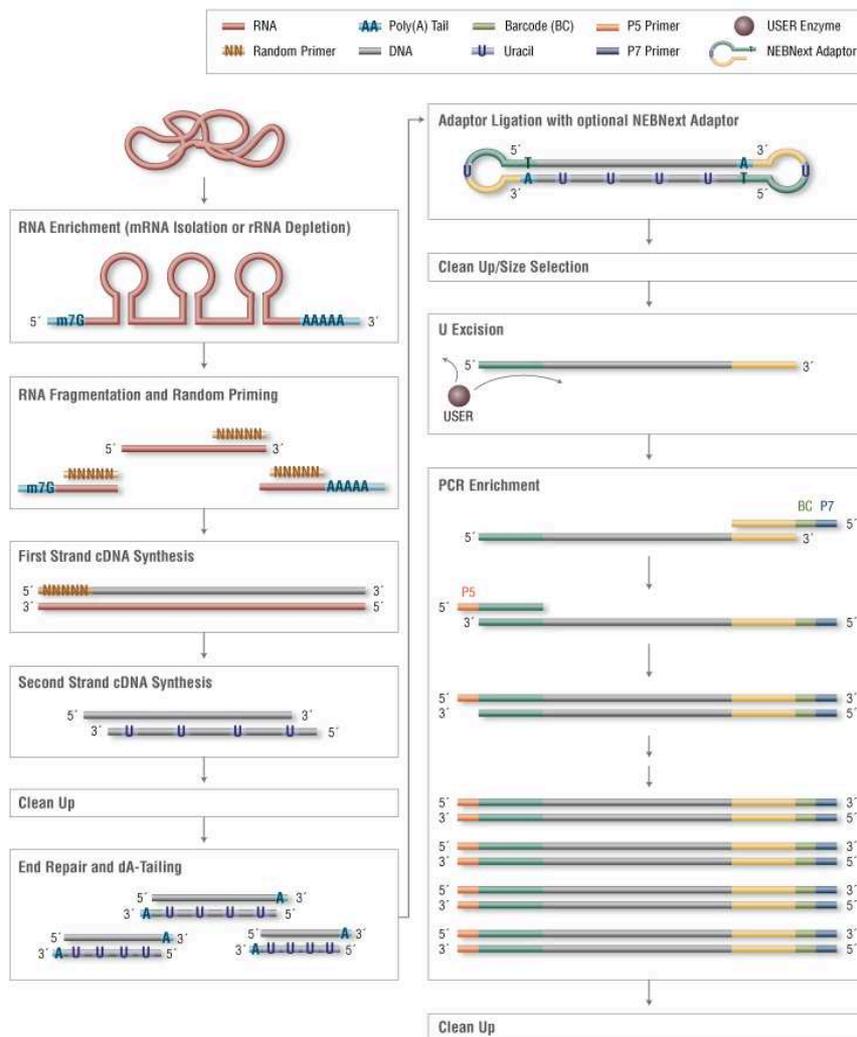


Figure 35 : Préparation des bibliothèques en vue d'une analyse RNA-Seq

Dans cette méthode, l'ARN est tout d'abord fragmenté par incubation à 94°C 15 min suivie d'une étape de reverse transcription par l'hybridation d'amorces aléatoires afin de synthétiser le premier brin d'ADNc. Pour synthétiser le deuxième brin d'ADNc, cette méthode repose sur l'utilisation de dUTP (désoxyuridine triphosphate) au lieu du désoxythymidine triphosphate (dTTP), marquant le second brin d'ADNc qui va être ensuite dégradé après l'ajout des adaptateurs par l'uracile-ADN glycosylase (UDG) afin de préserver l'orientation des transcrits. Ces adaptateurs permettent d'amplifier tous les milliards de fragments en même temps et de réaliser le séquençage à haut-débit ensuite. Enfin, une étape d'enrichissement par PCR et de multiplexage en ajoutant un index unique pour chaque échantillon en utilisant les oligos Multiplex NEBNext (NEB) a été effectuée.

❖ Analyse bioinformatique, statistique et contrôle qualité des données RNA-Seq

L'analyse des données de séquençage a été effectuée par Raphaël Lanos, ingénieur bioinformatien de notre unité, avec le pipeline automatisé AURIGA (**Figure 36**). Ce pipeline est composé de deux grandes parties, une partie bioinformatique et une partie statistique. Une troisième partie de contrôle qualité des données, indépendante du pipeline, a également été réalisée.

A- Analyse Bioinformatique

Les données brutes sorties du séquenceur sont démultiplexées avec le logiciel propriétaire illumina bcl2fastq2 v2.17, permettant ainsi d'obtenir des fichiers FASTQ contenant les séquences de chacun de nos échantillons. Ces séquences sont ensuite données en entrée à CutAdapt afin d'en retirer les morceaux de séquences des adaptateurs universels NEB. L'étape d'alignement des séquences en RNA-Seq est une étape critique car le logiciel doit être capable d'aligner des séquences ARN sur l'ADN en prédisant les jonctions d'épissage ou les points de fusion. Pour cette étape nous avons utilisé l'outil STAR v2.5.3a (Spliced Transcripts Alignment to a Reference, Dobin et al., 2013) avec les paramètres recommandés pour maximiser la précision et la spécificité de l'alignement sur le génome humain (Baruzzo et al., 2017). Les séquences ont donc été alignées sur le génome de référence GRCh37 release 75 d'ENSEMBL (<ftp://ftp.ensembl.org/pub/>) avec l'aide du fichier GTF (Gene Transfer Format) de cette même version comme guide d'annotation pour la prédiction des jonctions d'épissage. Le fichier GTF contient les coordonnées des exons et des transcrits des gènes codants mais aussi de pseudogène, de lincRNA, etc. Les résultats fournis au format SAM sont ensuite compressés et triés avec sambamba v0.6.6 (Tarasov et al., 2015) au format BAM. Enfin le comptage des séquences a été effectué avec l'outil featureCounts v1.5.2 (Liao et al., 2014) qui, avec l'aide du fichier GTF précédent, va compter pour chaque intervalle -dit exonique- toutes les séquences chevauchant cet intervalle. Les données sont ensuite agrégées au niveau du gène ou au niveau du transcrit afin de pouvoir observer les différences d'expression. L'utilisation des paramètres « -om -fraction » permet de gérer les cas de superposition des annotations, c'est à dire de comment attribuer les séquences aux gènes et aux transcrits lorsque celles-ci sont situées à un endroit du génome où plusieurs gènes/transcrits se chevauchent. Pour chacun de ces modes d'agrégation (gène ou transcrit), featureCounts fournit en plus un fichier de comptage de jonctions putatives. À la fin de cette étape d'analyse

AURIGA

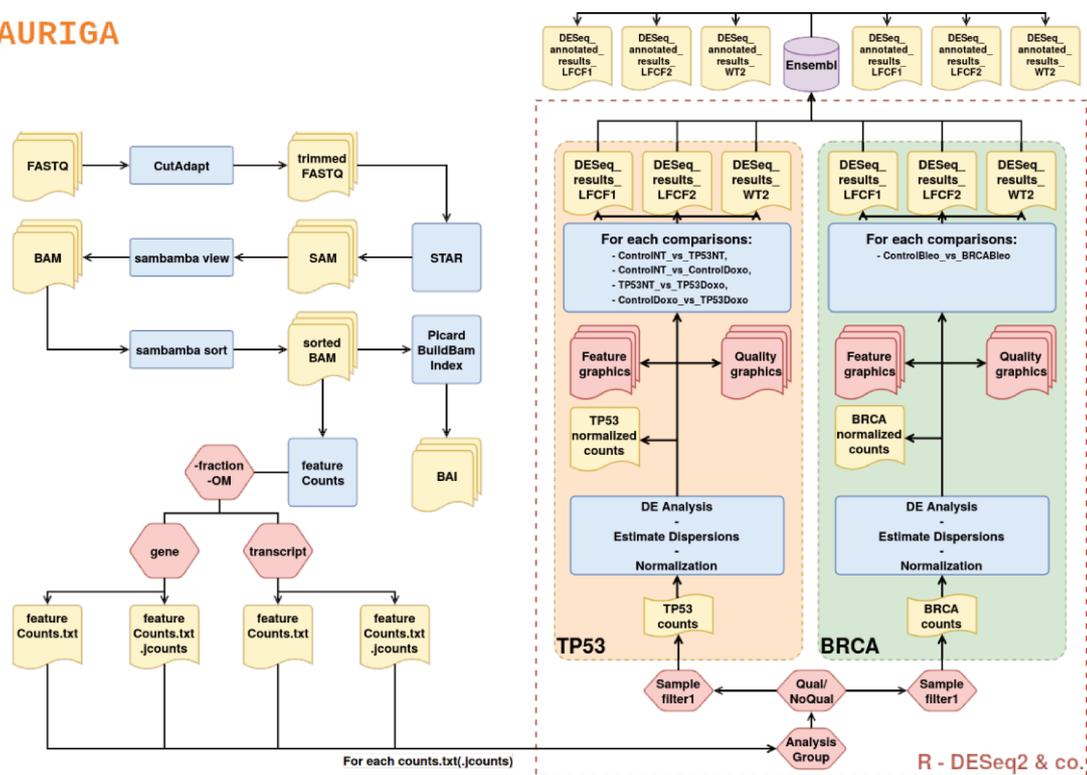


Figure 36 : Pipeline bioinformatique AURIGA dédié à l'analyse d'expression différentielle

bioinformatique nous avons donc pour chaque échantillon quatre fichiers de comptage, un fichier de comptage et un de jonctions putatives avec agrégation sur les gènes ainsi qu'un fichier de comptage classique et un de jonctions putatives avec agrégation sur les transcrits. Pour toutes nos analyses statistiques par la suite, nous avons travaillé sur les fichiers gènes.

B- Analyse Statistique

Dans le cadre de l'analyse statistique de l'expression différentielle des gènes ou des transcrits, les fichiers de comptage sont donnés en entrée à la librairie R (<https://www.r-project.org/>) DESeq2 v1.18.1 (Love et al., 2014). Les échantillons sont séparés en deux groupes BRCA et TP53 pour les analyses. Pour chacun de ces groupes, les données de comptage sont d'abord normalisées par DESeq2 puis filtrées selon un critère que nous avons défini en interne, à savoir : éliminer les gènes/transcrits pour lesquels nous n'avons pas au moins 20 séquences chez un échantillon. Le but de cette étape de filtration est d'éliminer le bruit de fond et de diminuer le nombre d'éléments à prendre en compte lors du test final de comparaison multiple. DESeq2 va ensuite faire l'estimation de la dispersion des données afin de donner des poids aux différents gènes/transcrits dans la définition du modèle statistique. DESeq2 va alors pouvoir modéliser les données en utilisant la distribution binomiale négative. Ensuite, en fonction de la question biologique posée, DESeq2 va calculer la p-value de chaque gène/transcrit en réalisant un test de Wald pour savoir si chacun d'entre eux est significativement différentiellement exprimé compte-tenu de la question posée. C'est à dire de savoir si le « log2FoldChange » est significativement différent de 0. Les questions qui ont été posées au modèle sont les suivantes :

1. Groupe TP53 :

- Sans traitement, quels sont les gènes/transcrits différentiellement exprimés entre les témoins et les mutants ?
- Avec traitement à la doxorubicine, quels sont les gènes/transcrits différentiellement exprimés entre les témoins et les mutants ?
- Chez les témoins, quels sont les gènes/transcrits différentiellement exprimés lors de l'application du traitement à la doxorubicine ?
- Chez les mutants, quels sont les gènes/transcrits différentiellement exprimés lors de l'application du traitement à la doxorubicine ?

2. Groupe BRCA :

- Quels sont les gènes/transcrits différentiellement exprimés entre les témoins traités à la bléomycine et les mutants traités à la bléomycine ?

Finalement, pour chaque question posée, le test ayant lieu sur de nombreuses entrées de gènes/transcrits, un test de correction multiple de type Benjamini-Hochberg est effectué afin d'obtenir la p-value finale, ajustée, de chacun des gènes/transcrits. Ensuite nous avons gardé les gènes significativement différentiellement exprimés ayant une p-value inférieure à 5 %. Les gènes/transcrits retenus étant sous forme d'identifiant ENSEMBL, une dernière étape d'analyse va annoter ces identifiants avec le nom et la description des gènes à partir de la base de données ENSEMBL.

C- Contrôle Qualité

Tout au long des analyses, différents contrôles de la qualité des données, utilisant divers outils, ont été effectués. Le contrôle qualité du séquençage et des séquences a été réalisé grâce à l'outil interne PyQua (<https://gitlab.bioinfo-diag.fr/NC4GPM/PyQua>). Le contrôle de la qualité de l'alignement a été fait à partir des données fournies par STAR. L'étape de comptage a été contrôlée avec des données de featureCounts. Avant de réaliser l'étape d'analyse statistique nous avons effectué un contrôle de la qualité de l'expérimentation en général en réalisant d'une part des analyses en composante principale et d'autre part un clustering hiérarchique avec l'aide respectivement, des bibliothèques R DESeq2 et pheatmap.

4- Développement de nouveaux essais fonctionnels

❖ Essai fonctionnel TP53 sans traitement génotoxique

Sur la base des données RNA-Seq sans traitement 6 gènes différentiellement exprimés entre le groupe des témoins et le groupe des mutants TP53 ont été sélectionnés (PTPRG [MIM * 176886 ; NM_002841], RNF180 [MIM * 616015 ; NM_0001113561], TLN2 [MIM * 607349 ; NM_015059], MGST1 [MIM * 138330 ; NM_001260511], SPATA18 [MIM * 612814 ; NM_145263], PDE3B [MIM * 602047 ; NM_000922]). Un amplicon de RT-QMPSF a été sélectionné pour chacun d'eux.

Ensuite une RT-QMPSF a été réalisée pour mesurer le taux d'expression de ces 6 gènes en utilisant un amplicon témoin (TBP [MIM * 600075, NM_003194]) pour normaliser les données. Les amplicons fluorescents ont été séparés par taille sur un analyseur génétique ABI Prism 3500 (Applied Biosystems, Foster City, CA) et les profils fluorescents résultants ont été analysés à l'aide du logiciel GeneScan 3.7 (Applied Biosystems) comme précédemment décrit. Les profils de RT-QMPSF des individus mutés pour p53 ont été comparés à ceux obtenus pour des témoins.

L'expression de ces gènes a été évaluée sur les ARN utilisés pour le RNA-Seq mais également sur 30 lignées immortalisées établies à partir de patients Li-Fraumeni porteurs de mutation hétérozygote du gène TP53.

❖ Essai fonctionnel TP53 sur prélèvement sanguin avec traitement génotoxique

Les échantillons sanguins ont été collectés dans des tubes de collecte de sang avec anticoagulant EDTA (acide Ethylène-Diamine-Tétra-Acétique) sous vide Vacutest®. L'isolement des cellules mononucléées du sang périphérique est effectué par une technique manuelle utilisant un gradient de Ficoll à différents temps après prélèvement allant de J0 à J+3. Après centrifugation, les cellules mononucléées se rassemblent en un anneau opaque surnageant au-dessus du Ficoll. Celui-ci est ensuite prélevé délicatement, lavé au PBS 3 fois dans le but d'éliminer une partie des thrombocytes contaminants et les débris cellulaires restés en suspension lors de la centrifugation et ensuite mis en suspension dans le milieu commercial Chromosome Medium P de chez AmpliTech (Compiègne, France) contenant de la phytohémagglutinine forme P, 10% de SVF, une interleukine, de l'héparine, de l'HEPES, et de la gentamicine comme antibiotique, afin de stimuler les lymphocytes et les faire entrer en division. Le comptage des cellules ainsi que la viabilité cellulaire ont été effectués sur le

compteur automatique des cellules NANOENTEK Adam avec le kit Kit AccuChip (SCIENCETEC, Villebon-sur-Yvette, France). Un million de cellules ont étéensemencées dans des plaques de 24 puits. Au moins 4 puits par individu sontensemencés pour avoir un puits non-traité et un puits traité en duplicats. 48 heures après la mise en culture les cellules sont traitées par 800 ng/ μ L de doxorubicine pendant 8h. Les cellules sont ensuite récoltées et centrifugées et l'extraction de l'ARN total est réalisée avec le kit NucleoSpin® RNA XS (Macherey Nagel) selon les instructions et quantifiés par spectrophotométrie à l'aide d'un spectrophotomètre UV-VIS ND-1000 (Biocompare, Nanodrop Technologies, USA). L'ARN a été congelé et stocké à -80°C immédiatement après la production.

❖ RT-MLPA

Sondes de RT-MLPA :

Cette réaction a été réalisée pour mesurer l'induction de la transcription de 10 gènes cibles de p53 sélectionnés sur la base des données de RNA-Seq (CEP170B ou KIAA0284 [NM_015005], PODXL [MIM * 602632, NM_001018111], RRAD [MIM * 179503, NM_004165], GLS2 [MIM * 606365, NM_013267], CABYR [MIM * 612135, NM_012189], TP53I3 [MIM * 605171, NM_004881]) EPS8L2 [MIM * 614988, NM_022772], SULF2 [MIM * 610013, NM_001161841]), SESN1 [MIM * 606103, NM_014454], FHL2 [MIM * 602633, NM_201555], et en utilisant trois amplicons témoins (TBP [MIM * 600075, NM_003194], RIC8B [MIM * 609147, NM_001330145] et MPP5 [MIM *606958, NM_022474.3) pour normaliser les données et deux sondes pour quantifier le niveau de transcrit du gène TP53 [MIM * 191170, NM_000546].

Les sondes de RT-MLPA ont été conçues sur les jonctions exon-exon pour éviter une amplification indésirable de l'ADN génomique. Toutes les sondes comprennent une région spécifique de l'ADNc cible. Les sondes 5' ont une queue additionnelle GTGCCAGCAAGATCCAATCTAGA à leur extrémité 5' et toutes les sondes 3' ont une queue additionnelle TCCAACCCTTAGGGAACCC à leur extrémité 3' pour permettre l'amplification finale par PCR. Des espaceurs de longueurs variables ont été insérés entre ces séquences et les séquences spécifiques à la cible pour permettre de distinguer les différents produits de PCR en fonction de leur taille. Toutes les sondes 3' ont été phosphorylées à leur extrémité 5' pour permettre les réactions de ligation. Des concurrents identiques à l'une des sondes mais sans les amorces ont été ajoutés au mélange de sondes RT-MLPA pour

normaliser les signaux d'amplification (en incluant des oligos concurrents non amplifiables dans un mélange de sondes MLPA, les signaux élevés peuvent être atténués à un niveau souhaité). Le mélange de sondes RT-MLPA a été préparé à partir de dilutions à 10 $\mu\text{mol/L}$ de sondes et de concurrents dans du Tris 10 mmol/L , et du tampon EDTA 1 mmol/L . Enfin toutes les sondes ont été réunies pour constituer un pool avec une concentration finale de 1 $\text{fmole}/\mu\text{L}$ chacune dans 10 mM de Tris / 1 mM d'EDTA pour créer le mix de RT-MLPA. Les séquences des sondes RT-MLPA sont fournies dans **Tableau Annexe 1**.

RT MLPA :

La reverse transcription a été faite selon les instructions du kit de synthèse d'ADN complémentaire Verso (Thermo scientific, Waltham, United States). 1 μL d'ARN total provenant de nos lignées correspondant à 100 ng, ont été ajoutés à 19 μL de mélange de transcription inverse du kit Verso. Les échantillons ont été incubés pendant 30 minutes à 42°C puis 2 minutes à 95°C. 6,25 μL du produit de reverse transcription ont été ajoutés à 1,5 μL du mélange de sonde de dilution finale et à 1,5 μL de tampon SALSA-MLPA (MRC-Holland, Amsterdam, Pays-Bas) avant la dénaturation à 95°C pendant 2 minutes et l'hybridation à 60°C pendant 1 heure. La ligation des sondes hybridées annelées a été effectuée à 54°C pendant 15 minutes, en ajoutant 32 μL de mélange de ligation, et chauffée pendant 5 minutes à 98 ° C. Ensuite, 2,5 μL du mélange de ligation ont été ajoutés à 7,5 μL d'un mélange de Salsa PCR contenant les amorces de PCR universelle dont l'une est marquée avec un 6-Fam. L'amplification par PCR impliquait 35 cycles de 94°C pendant 30 secondes, 58°C pendant 30 secondes et 72°C pendant 30 secondes, suivis de 72°C pendant 4 minutes.

Les amplicons de RT-MLPA résultants ont été analysés par analyse de fragments en utilisant un système d'électrophorèse capillaire ABI 3500 (Applied Biosystems, Foster City, CA). Pour l'analyse des fragments, 0,5 μL d'amplicon de PCR a été mélangé avec 19 μL de Hi-di Formamide et 0,5 μL de standard de taille Genescan-400 HD ROX (Applied Biosystems). Le mélange a été incubé pendant 2 minutes à 95°C et analysé. Les paramètres de migration étaient les suivants : Tension de pré-exécution, 15 kV ; Pre Run Time, 18 secondes ; Tension d'injection, 1,2 kV ; Temps d'injection, 5 secondes (réduit à 2 secondes lorsque les signaux de fluorescence étaient hors échelle), tension de fonctionnement, 10 kV ; Temps d'exécution, 900 secondes. Les amplicons fluorescents ont été séparés par taille et les profils fluorescents résultants ont été analysés à l'aide du logiciel GeneScan 3.7 (Applied

Biosystems). Les profils de RT-MLPA de cellules exposées et non exposées ont été superposés après ajustement des amplicons du contrôle à la même hauteur.

La hauteur du pic de chaque gène cible des 10 avec traitement à la doxorubicine était divisée par la somme des hauteurs des 3 gènes témoins. La résultante est divisée par le ratio de la hauteur du pic du gène cible sans traitement divisé par les hauteurs des 3 gènes contrôles. Le score arbitraire de fonctionnalité de p53 correspond à la moyenne des scores obtenus pour chacun des 10 gènes cibles.

Pour vérifier les niveaux d'ARNm de p53 dans nos lignées mutées, la hauteur de chacun des pics a été divisée par la somme des hauteurs des pics contrôles au niveau de la lignée mutante TP53. Le ratio obtenu a été comparé à celui obtenu chez les témoins.

❖ RT-QMPSF

Une réaction de RT-QMPSF a été réalisée selon le protocole décrit ci-dessus pour mesurer l'induction de la transcription des 10 gènes cibles sélectionnés en RT-MLPA (KIAA0284, PODXL, RRAD, GLS2, CABYR, TP53I3, EPS8L2, SULF2, SESN1, FHL2) dans les lignées mutées pour TP53. Les mêmes gènes témoins (TBP, RIC8B et MPP5) ont été utilisés pour normaliser les données. Deux amplicons ont été choisis pour quantifier le niveau de transcrit du gène TP53. La méthode de calcul des scores de fonctionnalité appliquée est la même que celle décrite dans le paragraphe RT-MLPA.

III- Résultats

1- Induction d'un stress génotoxique

Nous avons choisi d'évaluer la pertinence des méthodes multi-omiques pour différencier une cellule normale d'une cellule mutée à l'état hétérozygote. Nous avons pour cela utilisé des lignées lymphoblastoïdes dérivées de patients ou d'individus témoins indemnes de cancer. Pour ce projet, nous avons sélectionné 4 lignées établies à partir de patients Li-Fraumeni porteurs de mutations délétères à l'état hétérozygote du gène TP53 : - 3 lignées avec des mutations faux-sens à effet dominant négatif [NM_000546.4 : p.(Arg175His), p.(Arg248Trp), p.(Arg273His)] et - une mutation « nulle » (NM_000546.4 : délétion totale). Nous avons à disposition 2 lignées avec mutation délétère à l'état hétérozygote du gène BRCA1 (NM_007294.3 : c.81-80_547+81dup; NM_007294.3 : c.1-?_80+?del) et 2 lignées avec mutation délétère à l'état hétérozygote du gène BRCA2 (NM_000059.3: c.5993_5994del ; NM_000059.3: c.7008-1G>A). Ces lignées ont été comparées à 4 lignées EBV dérivées de témoins indemnes de cancer (T401, T403, T404, T-SK).

En nous inspirant de l'essai fonctionnel développé par notre équipe pour TP53 décrit dans la première partie des résultats, nous avons cherché à **démasquer l'endophénotype des cellules mutées** en générant un **stress génotoxique**. Nous avons donc traité nos cellules avec des agents induisant des lésions au niveau de l'ADN. Pour les mutations du gène TP53 impliqué dans la réponse aux lésions de l'ADN, nous avons appliqué la **doxorubicine**, un inhibiteur d'ADN Topoisomerase II de la famille des anthracyclines utilisé comme chimiothérapie. Pour les cellules mutées pour les gènes BRCA1 ou BRCA2 impliqués dans la réparation des cassures double brin de l'ADN par recombinaison homologue, nous avons choisi d'utiliser la **bléomycine** un antibiotique anti-cancéreux induisant des cassures simple et double brin de l'ADN. Nous avons établi les meilleures conditions de traitement (concentration de la drogue et durée de traitement) à l'aide de l'essai fonctionnel TP53, décrit dans la première partie, utilisé comme test de génotoxicité (Zerdoumi et al., 2015). La dose retenue pour la doxorubicine était de 200 ng/ml soit 0,3 μ M et celle pour la bléomycine était de 10 μ g/mL soit 7 μ M pour une durée de 8h de traitement.

Pour déterminer si le traitement avait réellement eu l'effet génotoxique escompté pour tous les échantillons, nous avons réalisé cet essai sur chacun des échantillons générés en triplicats biologiques. Les scores obtenus pour toutes les lignées figurent dans le **Tableau 5**. Ces scores montrent que tous les gènes cibles de p53 réagissent au stress chez les témoins et

chez les mutants suggérant l'activation appropriée des voies de réponse à la drogue après notre traitement.

Scores et traitements		Lignées				Lignées TP53 WT/MT				Lignées BRCA WT/MT			
		T401	T403	T404	T-SK	R248W	R273H	R175H	Del-totale	1823	827	1820	1838
Scores d'induction	Doxorubicine 0.3 μ M, 8h	11,7	9,72	10,49	11,68	1,64	2,03	4,31	4,92	-	-	-	-
	Bléomycine 7 μ M, 8h	5,42	5,7	4,24	4,27	-	-	-	-	4	2,20	4,26	3,43

Tableau 5 : Scores d'induction de p53 obtenus en test fonctionnel de RT-QMPSF après traitement avec la doxorubicine ou la bléomycine pour toutes les lignées analysées en RNA-Seq

WT: Wild Type; MT: Mutant

2- Analyse non ciblée du métabolome

Pour détecter les différences potentielles existant entre des cellules de génotype sauvage et des cellules de patients mutées à l'état hétérozygote, nous avons choisi une approche de métabolomique qui représente l'analyse de tous les métabolites présents à un instant T dans un échantillon biologique. Cette partie de ma thèse a été réalisée en collaboration avec l'équipe de Biochimie Métabolique du CHU de Rouen dirigée par le Pr Soumeya Bekri et l'unité UMR 6014 Laboratoire de Chimie Organique et Analytique COBRA dirigée par le Pr Carlos Afonso.

Nous avons réalisé les cultures cellulaires en triplicats biologiques pour les lignées témoins et les lignées issues de patients mutés pour *TP53*, *BRCA1* ou *BRCA2* avec ou sans traitement génotoxique. Après les huit heures de traitement, les culots cellulaires ont été récupérés pour extraction des métabolites. La méthode d'extraction que nous avons utilisé (précipitation des protéines au méthanol après lyse cellulaire) vise à conserver la plus grande partie des métabolites polaires mais ne permet pas de conserver les métabolites lipidiques. Après extraction, les échantillons ont été analysés par **spectrométrie de masse à haute résolution couplée à la mobilité ionique et à la chromatographie liquide ultra-haute performance** au laboratoire de Chimie Organique et Analytique COBRA par le Dr Abdellah Tebani (Unité de Biochimie Métabolique, CHU de Rouen). Les données ont ensuite été filtrées pour éliminer le bruit de fond, les pics sont détectés et un alignement est réalisé entre les différents profils. Les données sont alors normalisées, transformées et analysées de façon statistique. Cette analyse statistique a été réalisée par le Dr Abdellah Tebani.

Une analyse en composante principale a été réalisée pour visualiser la tendance au regroupement des échantillons mutés ou sauvages. Cette analyse non supervisée n'a pas permis de distinguer des groupes clairement définis (données non présentées). Dans le but de créer un modèle prédictif de la présence d'une mutation dans le gène *TP53* ou dans l'un des gènes *BRCA*, l'outil mathématique OPLS-DA (*Orthogonal Partial Least Square Discriminant Analysis*) a été appliqué à nos données. Dans cette analyse supervisée, on définit un groupe d'échantillons témoins et un groupe d'échantillons mutés. Un modèle est alors créé en retenant la combinaison de métabolites la plus discriminante pour différencier les deux groupes. Ce modèle sera par la suite appliqué à chaque nouvel échantillon testé pour déterminer si, sur la base des variables retenues pour le modèle, l'échantillon se rapproche plus du groupe des témoins non mutés ou du groupe des patients avec mutation délétère. Cette analyse réalisée pour les lignées avec mutation délétère de *TP53* versus les lignées témoins ne permet pas de façon significative de différencier les deux groupes en conditions basales (sans traitement) (p-value = 1) ni avec le traitement génotoxique (doxorubicine) (p-value = 1) (**Figure 37**).

De même pour les lignées mutées dans l'un des gènes *BRCA*, sans traitement génotoxique nous ne sommes pas capables de différencier les groupes mutés et témoins de façon significative dans les conditions basales (p-value = 0,13) et dans les conditions où un stress génotoxique est appliqué (bléomycine 7 μ M pendant 8h) (p-value = 0,13) (**Figure 38**). Il est à noter que pour les lignées *BRCA*, les échantillons ont été générés en triplicats biologiques mais n'ont pas pu être tous passés sur le spectromètre de masse. Les résultats présentés dans la **Figure 38** représentent l'analyse de duplicats et comportent 4 lignées témoins additionnelles.

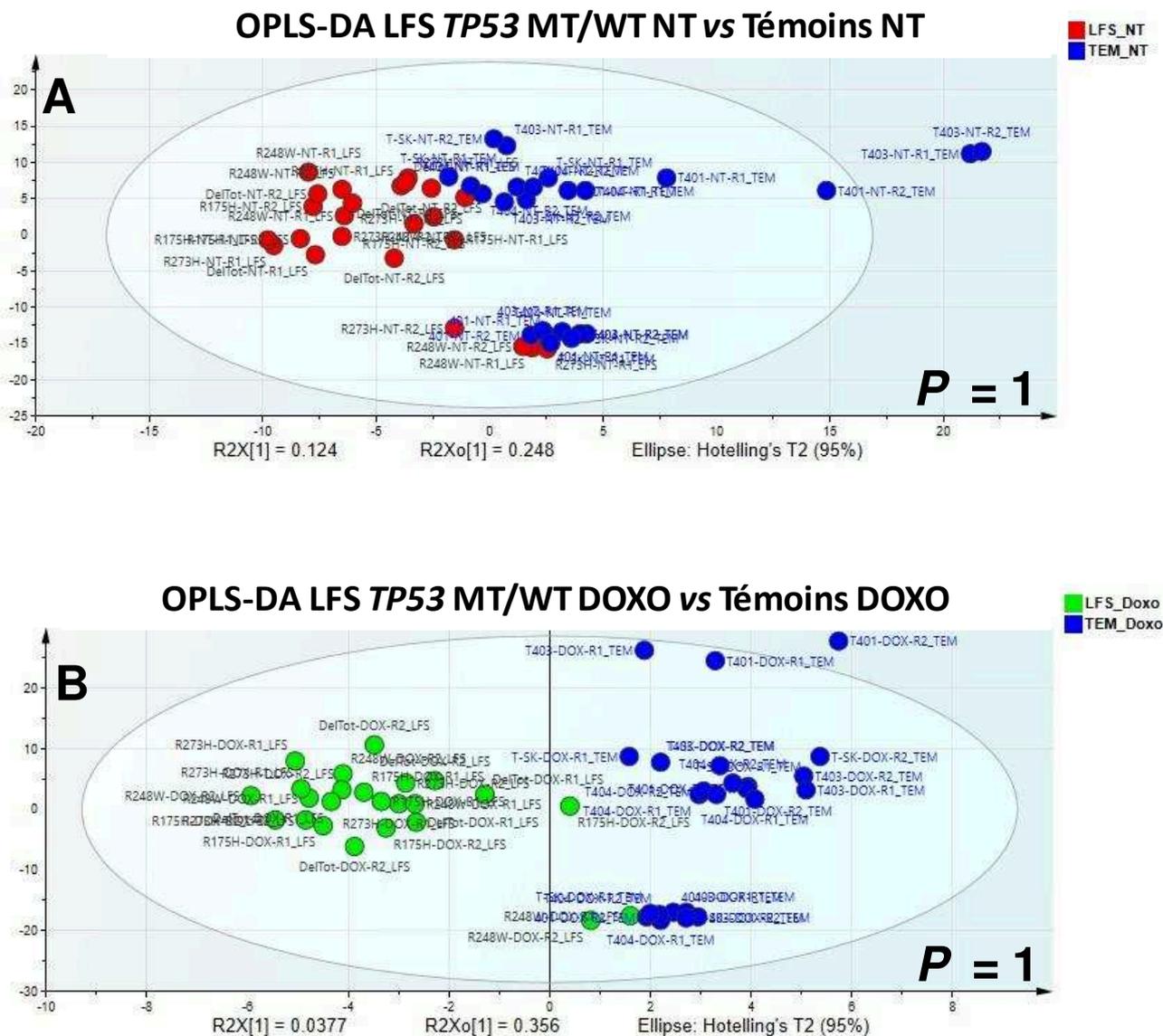


Figure 37 : Analyse OPLS-DA des données de métabolomique obtenues pour les lignées mutées pour *TP53* et les lignées témoins

- A- Séparation des échantillons témoins (en bleu) et des mutés *TP53* (en rouge) en condition basale (NT, sans traitement)
- B- Séparation des échantillons témoins (en bleu) et des mutés *TP53* (en vert) avec traitement génotoxique (doxorubicine 0,3 μ M pendant 8h)

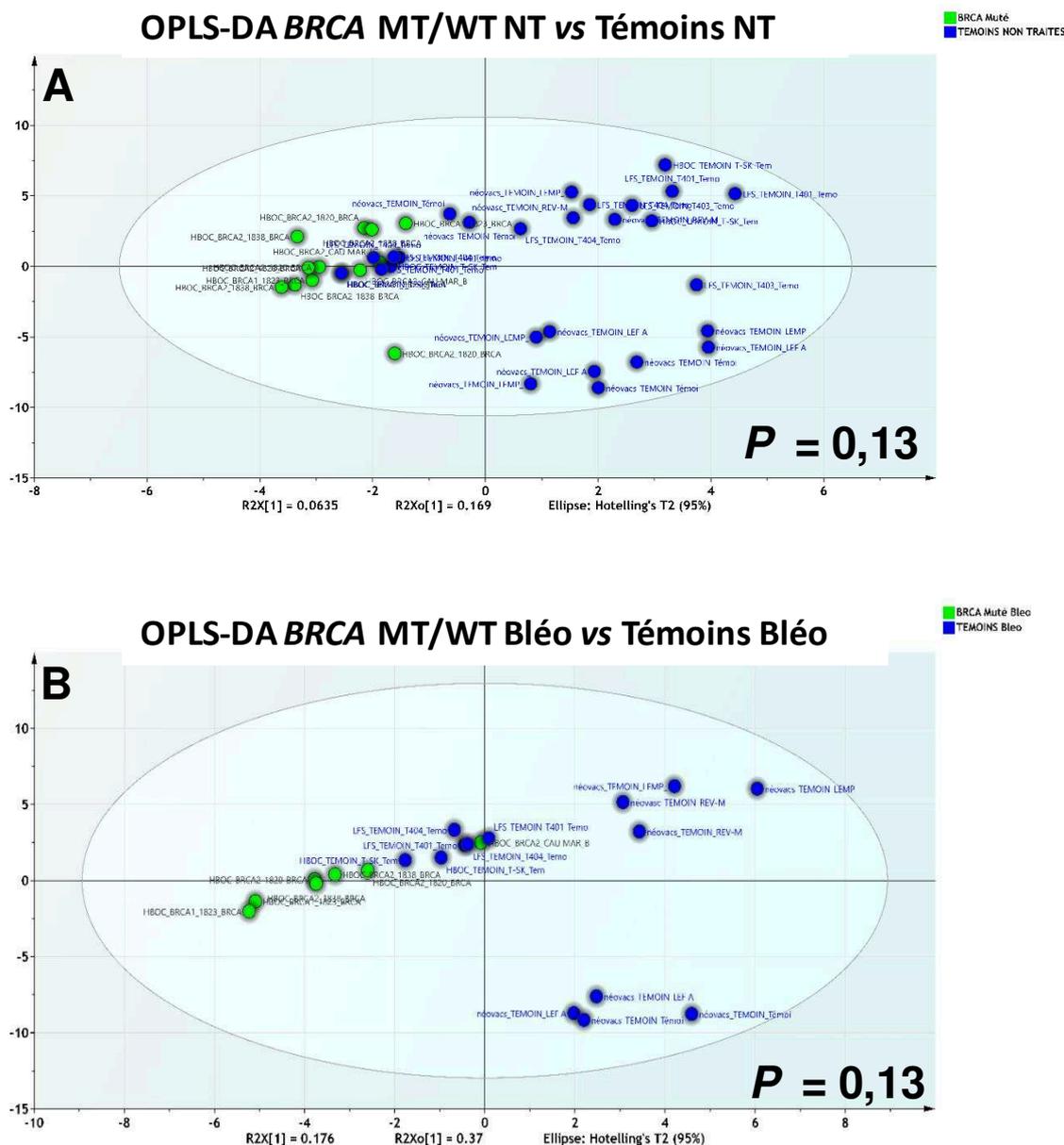


Figure 38 : Analyse OPLS-DA des données de métabomomique obtenues pour les lignées mutées pour l'un des gènes *BRCA* et les lignées témoins

- A-** Séparation des échantillons témoins (en bleu) et des mutés pour l'un des gènes *BRCA* (en vert) en condition basale (NT, sans traitement)
- B-** Séparation des échantillons témoins (en bleu) et des mutés pour l'un des gènes *BRCA* (en vert) avec traitement génotoxique (bléomycine 7 μ M pendant 8h)

3- Analyse du transcriptome des cellules sauvages et mutées par RNA-Seq total

Comme deuxième approche, nous avons choisi une analyse de **transcriptomique**. Nous avons effectué une analyse de **RNA-seq total en triplicats biologiques**. Pour cette analyse, nous avons utilisé les 4 lignées témoins, les 4 lignées mutées *TP53* à l'état hétérozygote et les 4 lignées mutées pour l'un des gènes *BRCA*. Nous avons traité nos cellules avec les agents génotoxiques pendant 8 heures avant de récupérer les cellules pour réaliser une extraction des longs ARN. Nous avons eu la possibilité d'inclure dans les échantillons analysés en RNA-Seq les lignées mutées *TP53* à l'état basal que nous avons pu comparer aux lignées sauvages non traitées. Nous avons procédé à une déplétion des ARN ribosomaux qui représentent près de 80 % du transcriptome humain. Ensuite, nous avons fragmenté l'ARN, effectué une reverse transcription à l'aide d'hexamères aléatoires et nous avons finalement ligué les adaptateurs qui permettent d'amplifier tous les milliards de fragments en même temps et de réaliser le séquençage à haut-débit ensuite.

Le séquençage a été réalisé sur le séquenceur Illumina Nextseq500 de notre unité pour générer en moyenne 50 millions de reads pairés par échantillon (**Tableau 6**). En moyenne plus de 93 % des bases ont été lues avec un Qscore supérieur à 30 et la qualité moyenne de séquençage était de 34 (**Tableau 6**). Les lectures ont ensuite été alignées sur le génome de référence humain hg19. Environ 85 % des lectures ont pu être alignées. Le contrôle qualité de l'alignement a été fait à partir des données fournies par l'outil aligneur STAR et nous a permis de détecter que l'échantillon 827-Bleo-3 avait un taux anormalement élevé de séquences alignées à de multiples endroits (**Figure 7**). Ce point de réplicat, « outlier », a donc été par la suite éliminé de l'analyse. L'analyse de RNA-seq nous a permis de détecter environ 118 000 transcrits correspondant à 22720 gènes par échantillon. A l'aide du pipeline d'analyse AURIGA, développé par Raphaël Lanos au laboratoire, automatisant les outils STAR, Feature Counts et DESeq2, nous avons procédé à une analyse d'expression différentielle sur les données *TP53* d'une part, puis sur les données *BRCA*. Ensuite, Abdellah Tebani a appliqué l'outil mathématique OPLS-DA sur les données de RNA-Seq pour établir un modèle prédictif de la présence d'une mutation délétère de *TP53* ou d'un des gènes *BRCA*.

Tableau 6 : Contrôle qualité des données de RNA-Seq

Echantillon	Nombre moyen de paires de reads	% \geq Q30 bases	Moyenne du score de qualité (Qscore)	Echantillon retenu
BRCA1 -1823-Bleo-R1	68385702,0	94,22	34,41	OK
BRCA1 -1823-Bleo-R2	61564196,8	93,46	34,25	OK
BRCA1 -1823-Bleo-R3	69923008,0	94,28	34,43	OK
BRCA1 -827-Bleo-R1	58479373,0	92,935	34,16	OK
BRCA1 -827-Bleo-R2	67441006,0	94,06	34,38	OK
BRCA1 -827-Bleo-R3	73411136,8	91,40	33,84	OK
BRCA2 -1838-Bleo-R1	59322547,0	91,25	33,80	OK
BRCA2 -1838-Bleo-R2	62747305,0	94,08	34,39	OK
BRCA2 -1838-Bleo-R3	63071359,0	94,24	34,42	OK
BRCA2 -1820-Bleo-R1	63275423,0	93,99	34,37	OK
BRCA2 -1820-Bleo-R2	58219866,0	92,75	34,12	OK
BRCA2 -1820-Bleo-R3	53354703,0	92,47	34,06	OK
Témoin-T401-Bleo-R1	59249847,0	93,60	34,29	OK
Témoin-T401-Bleo-R2	66515600,0	93,88	34,35	OK
Témoin-T401-Bleo-R3	61561632,0	94,11	34,40	OK
Témoin-T403-Bleo-R1	70814493,0	94,05	34,39	OK
Témoin-T403-Bleo-R2	70228574,0	93,93	34,36	OK
Témoin-T403-Bleo-R3	66455512,0	92,85	34,12	OK
Témoin-T404-Bleo-R1	57873548,0	92,79	34,13	OK
Témoin-T404-Bleo-R2	71402653,0	94,23	34,42	OK
Témoin-T404-Bleo-R3	78197960,0	94,15	34,40	OK
Témoin-T-SK-Bleo-R1	66148068,0	92,75	34,12	OK
Témoin-T-SK-Bleo-R2	57212010,0	90,87	33,71	OK
Témoin-T-SK-Bleo-R3	60579566,0	92,84	34,14	OK
TP53 -R248W-NT-R1	71258901,0	93,13	34,22	OK
TP53 -R248W-NT-R2	68935777,0	92,84	34,14	OK
TP53 -R248W-NT-R3	64808091,0	93,05	34,19	OK
TP53 -R273H-NT-R1	68701309,0	93,24	34,24	OK
TP53 -R273H-NT-R2	68087042,0	93,19	34,22	OK
TP53 -R273H-NT-R3	58368955,0	93,52	34,29	OK
TP53 -R175H-NT-R1	63534750,0	93,36	34,27	OK
TP53 -R175H-NT-R2	73184810,0	93,38	34,25	OK
TP53 -R175H-NT-R3	64276684,0	93,43	34,27	OK
TP53 -deltotale-NT-R1	70307805,0	93,41	34,28	OK
TP53 -deltotale-NT-R2	62613280,0	93,22	34,22	OK
TP53 -deltotale-NT-R3	71199639,0	93,45	34,28	OK

Objectif 2

Témoïn-T401-NT-R1	61846891,0	93,05	34,20	OK
Témoïn-T401-NT-R2	64402651,0	93,29	34,24	OK
Témoïn-T401-NT-R3	61853691,0	93,53	34,29	OK
Témoïn-T403-NT-R1	62957009,0	92,82	34,15	OK
Témoïn-T403-NT-R2	63341734,0	92,90	34,15	OK
Témoïn-T403-NT-R3	63886678,0	93,31	34,25	OK
Témoïn-T404-NT-R1	59765235,0	92,33	34,04	OK
Témoïn-T404-NT-R2	67860206,0	92,52	34,07	OK
Témoïn-T404-NT-R3	62101126,0	92,50	34,07	OK
Témoïn-T-SK-NT-R1	62418997,0	93,12	34,21	OK
Témoïn-T-SK-NT-R2	67670940,0	93,22	34,22	OK
Témoïn-T-SK-NT-R3	70845190,0	93,50	34,29	OK
TP53 -R248W-Doxo-R1	68469013,0	94,80	34,55	OK
TP53 -R248W-Doxo-R2	61827117,0	94,59	34,49	OK
TP53 -R248W-Doxo-R3	66874932,0	91,60	33,86	OK
TP53 -R273H-Doxo-R1	53092624,0	94,45	34,47	OK
TP53 -R273H-Doxo-R2	73076467,0	91,20	33,78	OK
TP53 -R273H-Doxo-R3	70146363,0	91,63	33,87	OK
TP53 -R175H-Doxo-R1	55570359,0	91,62	33,86	OK
TP53 -R175H-Doxo-R2	61769037,0	92,27	34,01	OK
TP53 -R175H-Doxo-R3	77527706,0	93,31	34,24	OK
TP53 -deltotale-Doxo-R1	65109201,0	92,81	34,13	OK
TP53 -deltotale-Doxo-R2	66222465,0	93,66	34,31	OK
TP53 -deltotale-Doxo-R3	72158253,0	93,21	34,21	OK
Témoïn-T401-Doxo-R1	60895710,0	94,71	34,52	OK
Témoïn-T401-Doxo-R2	68566111,0	93,37	34,25	OK
Témoïn-T401-Doxo-R3	69288108,0	91,27	33,79	OK
Témoïn-T403-Doxo-R1	59968009,0	94,64	34,51	OK
Témoïn-T403-Doxo-R2	56406423,0	94,51	34,48	OK
Témoïn-T403-Doxo-R3	64920485,0	89,81	33,48	OK
Témoïn-T404-Doxo-R1	57926450,0	94,02	34,37	OK
Témoïn-T404-Doxo-R2	75433566,0	91,73	33,90	OK
Témoïn-T404-Doxo-R3	71594394,0	93,48	34,28	OK
Témoïn-T-SK-Doxo-R1	66798182,0	94,64	34,50	OK
Témoïn-T-SK-Doxo-R2	69783943,0	93,27	34,23	OK
Témoïn-T-SK-Doxo-R3	66292480,0	91,69	33,88	OK

NT : Non traité ; Bleo : Bléomycine ; Doxo : Doxorubicine ; R1, 2, 3 : Réplicats 1, 2, 3

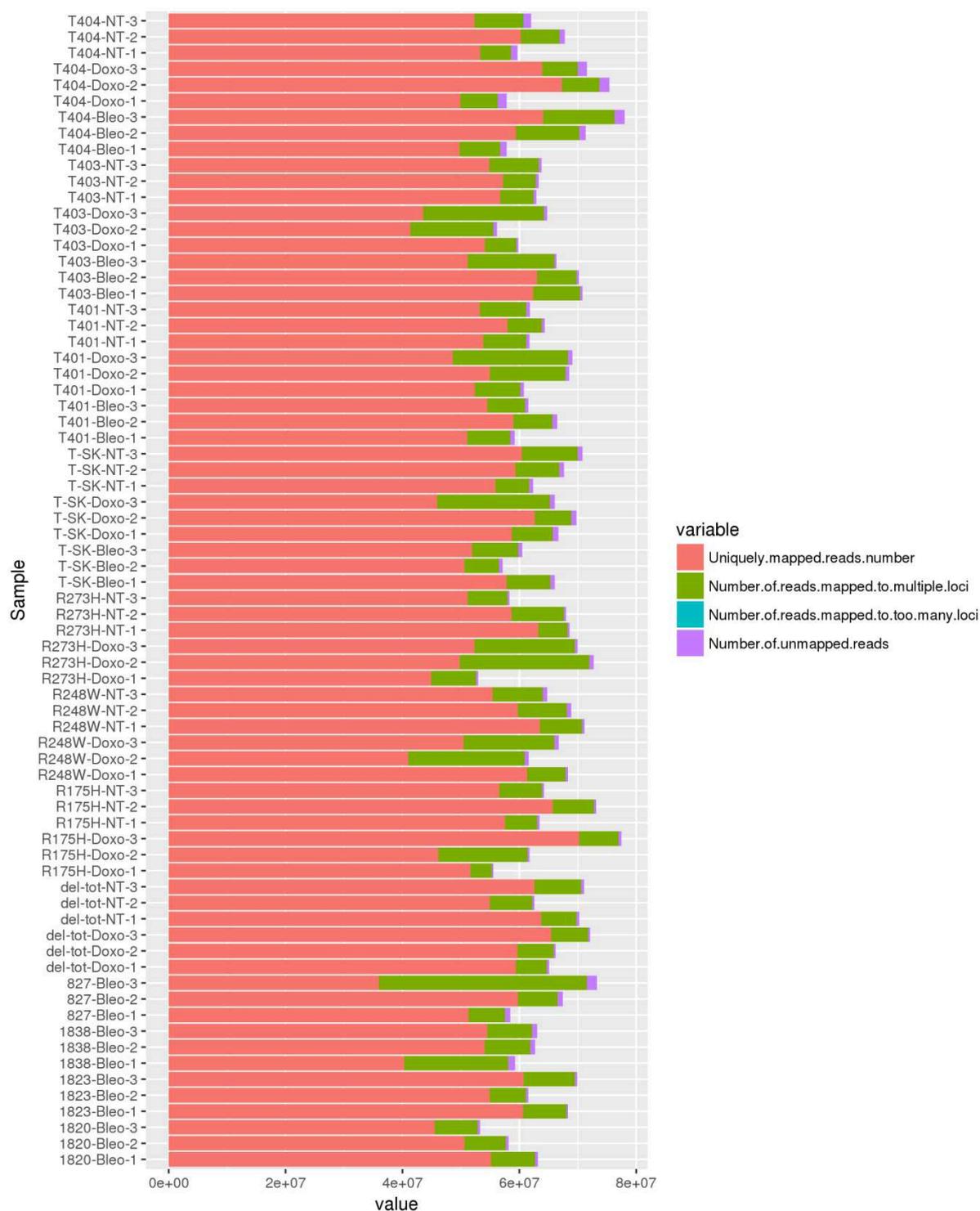


Figure 39 : Résultats de l’alignement sur le génome de référence humain hg19

Environ 85 % des lectures ont pu être alignées (en rose) sur des séquences uniques. Les reads alignés sur plusieurs endroits du génome sont en vert et ceux non mappés sont présentés en violet.

A- Comparasion des lignées TP53 Wt/Mt et des lignées témoins

Nous avons dans un premier temps comparé les transcriptomes des cellules sauvages et des cellules portant une mutation délétère de *TP53*. Nous avons eu l'opportunité de passer nos échantillons dans deux conditions :

- a- **Sans traitement** génotoxique, en espérant détecter une différence à l'état basal entre les cellules normales et les cellules mutées à l'état hétérozygote pour *TP53*.
- b- **Avec traitement** génotoxique (doxorubicine 0,3 μ M) pour exacerber les différences potentielles existant entre les cellules non mutées et mutées.

Avant de réaliser l'analyse d'expression différentielle, nous avons effectué une **analyse en composantes principales** d'une part et d'autre part un **clustering hiérarchique** avec l'aide respectivement, des librairies R DESeq2 et de Pheatmap. L'analyse en composante principale révèle, par la faible variance de ces deux premiers axes, que nos échantillons ne peuvent être catégorisés efficacement selon deux axes, leur différenciation étant complexe pour ce type de modèle (données non présentées). Le clustering hiérarchique présenté en **Figure 40** tend à rassembler les échantillons par leur génotype. Cette analyse non supervisée met en évidence que le premier critère de différenciation des échantillons est principalement la lignée cellulaire, puis le traitement.

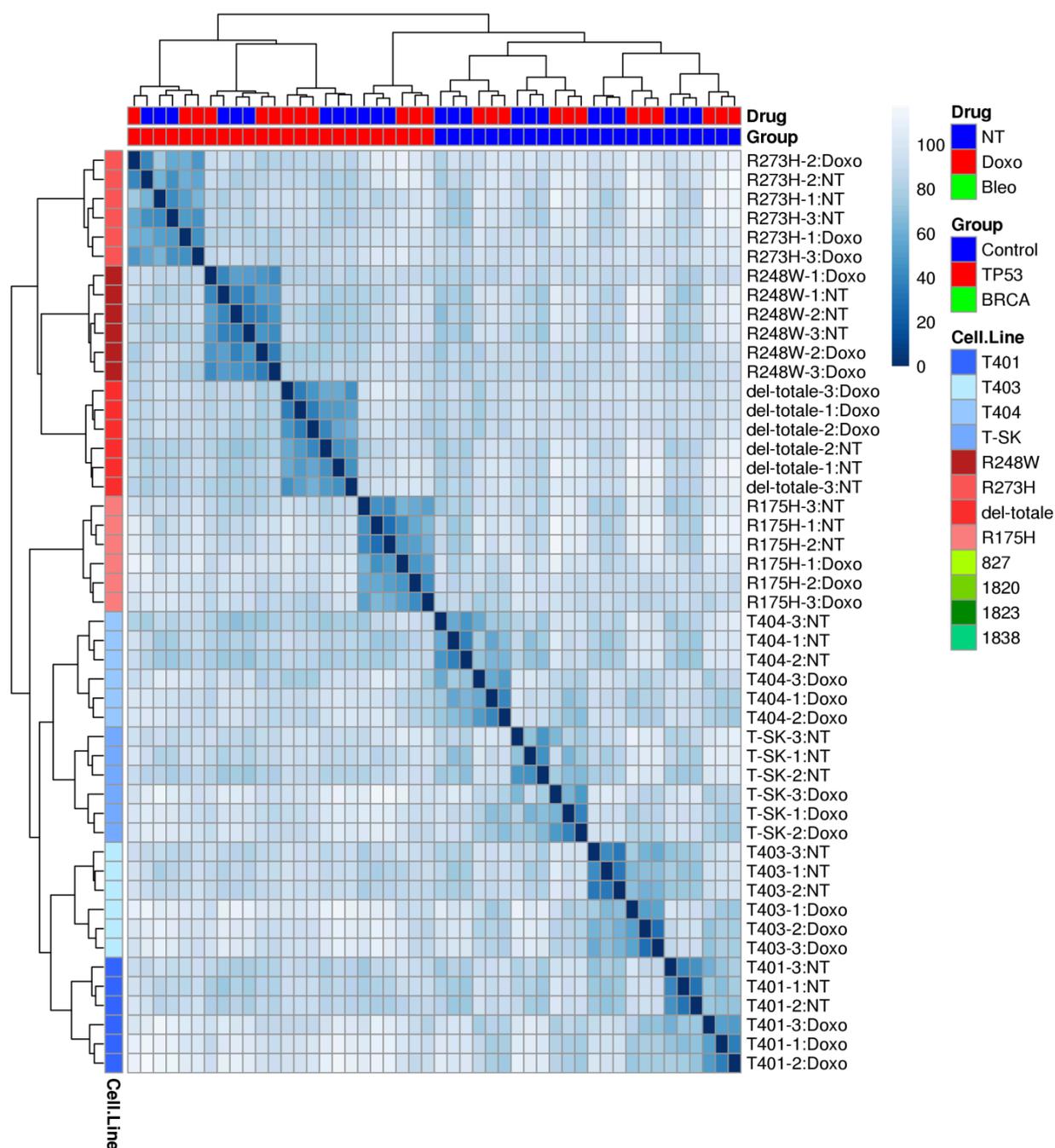


Figure 40 : Clustering hiérarchique (non supervisé) des données de RNA-Seq obtenues pour les lignées témoins et les lignées mutées *TP53* avec et sans traitement à la doxorubicine

NT : Non traité ; Doxo : Doxorubicine ; 1, 2,3 : Réplicats 1, 2, 3.

a- Comparaison des lignées à l'état basal**- Analyse d'expression différentielle**

Nous avons réalisé une analyse d'expression différentielle entre les lignées sauvages ou mutantes pour *TP53* à l'état basal sans traitement génotoxique. Cette analyse a été réalisée à l'aide de l'outil DESeq2 avec comme filtres une p-value inférieure à 5 % et un log2Fold change supérieur à 2 ou inférieur à -2 ce qui correspond à une sur-expression de 4 fois ou à une sous-expression de 4 fois. Avec ces critères, nous avons mis en évidence 444 gènes différentiellement exprimés entre le groupe témoin et le groupe des mutants *TP53*. Parmi ces gènes, un tiers correspond à des gènes d'ARN non-codants. 266 sont sous-exprimés chez les mutants par rapport aux contrôles alors que 178 sont sur-exprimés chez les mutants.

Afin d'aider à l'interprétation biologique des listes de gènes d'intérêt identifiés, nous avons cherché à regrouper les gènes par similarité fonctionnelle. Pour ce faire, nous avons utilisé le logiciel DAVID 6.8 gratuit et interrogeable en ligne (<https://david.ncifcrf.gov/home.jsp>). Reposant sur des milliers d'annotations fonctionnelles issues de la Gene Ontology (Processus biologiques, fonctions moléculaires, localisation cellulaire), mais également sur des bases de données de voies de signalisation (KEGG, Biocarta, Reactome,...), ce logiciel permet de déterminer si une liste de gènes d'intérêt présente un enrichissement fonctionnel par rapport au génome de référence. Cet enrichissement est déterminé à l'aide du score EASE (Expression Analysis Systematic Explorer) qui repose sur un test exact de Fisher unilatéral modifié. Cet outil de classification fonctionnelle offre ainsi un moyen rapide d'organiser de grandes listes de gènes en groupes fonctionnellement liés afin d'aider à démêler le contenu biologique capturé par les technologies à haut-débit (Huang et al., 2007).

L'analyse ontologique réalisée sur les 444 gènes différentiellement exprimés révèle une multitude de voies biologiques différentes parmi lesquelles la neurogenèse, la migration cellulaire ou l'organisation des membranes cellulaires (**Figure 41**).

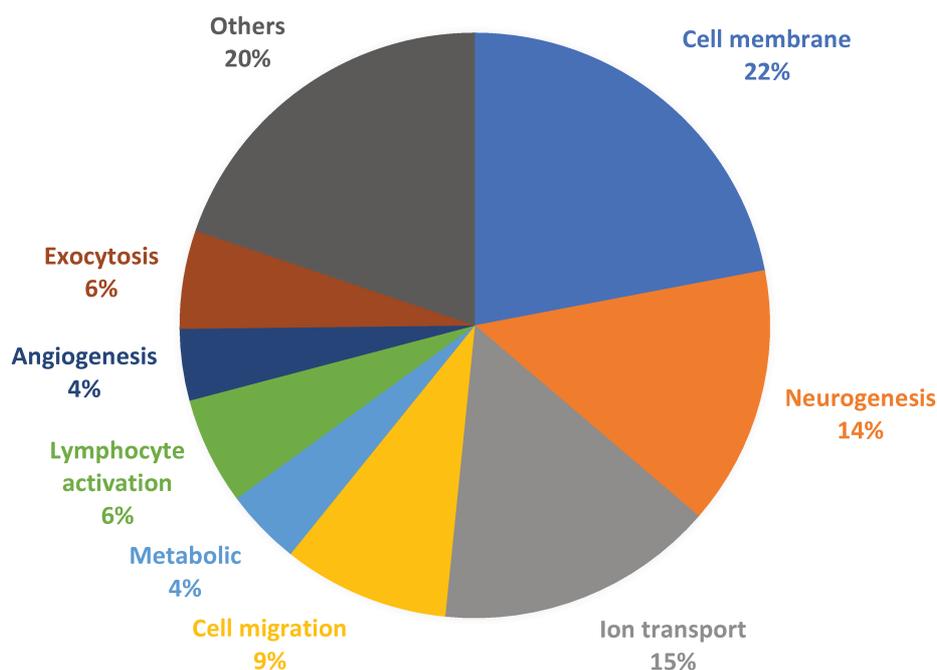


Figure 41 : Analyse d'Ontologie réalisée sur les gènes différentiellement exprimés en conditions basales entre les lignées mutées pour *TP53* et les lignées témoins

- **Création d'un modèle prédictif (analyse OPLS-DA)**

Lorsque nous réalisons une analyse comparative supervisée sur nos données de RNA-Seq avec l'outil OPLS-DA que nous avons appliqué pour l'analyse métabolomique, les groupes de témoins et de mutants pour *TP53* sont clairement différenciés avec une p-value très encourageante de l'ordre de 10^{-5} (**Figure 42**). Ces résultats très intéressants semblent indiquer qu'il serait possible de distinguer des cellules avec mutation délétère de *TP53* de cellules de génotype normal et ce même en l'absence d'un stress génotoxique. Ces résultats restent bien sûr à confirmer sur une plus grande série d'échantillons.

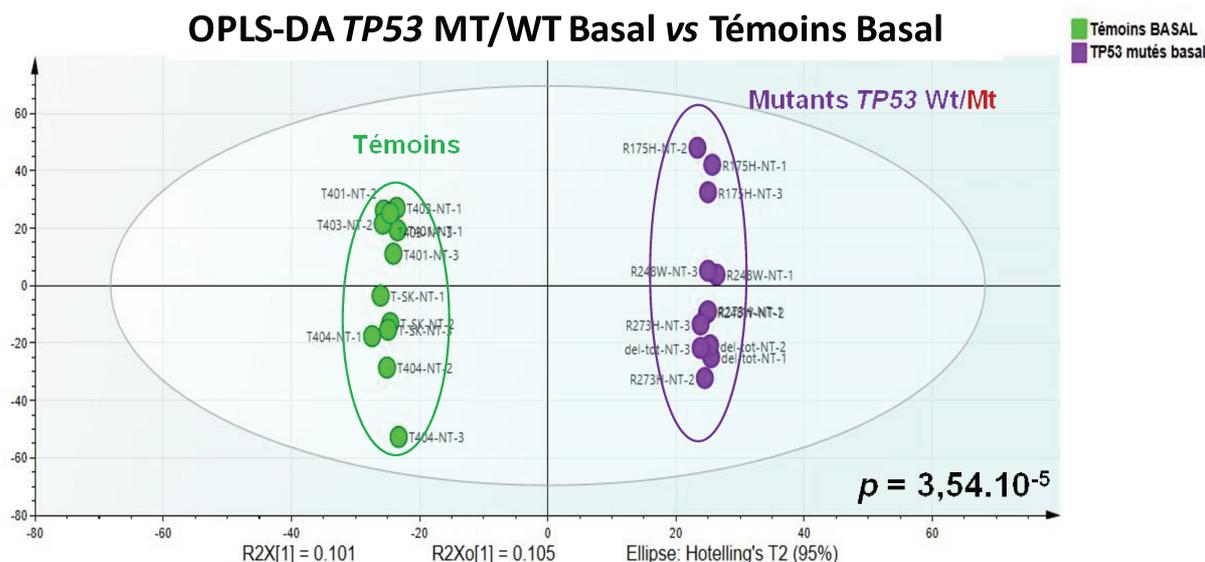


Figure 42 : Analyse OPLS-DA des données de RNA-Seq obtenues pour les lignées mutées pour le gène *TP53* et les lignées témoins à l'état basal

Séparation des échantillons témoins (en vert) et des mutés pour le gène *TP53* (en violet) en condition basale (sans traitement).

b- Comparaison des lignées en présence d'un traitement génotoxique

- Analyse d'expression différentielle

Nous avons réalisé une analyse d'expression différentielle entre les lignées sauvages ou mutantes pour *TP53* en présence d'un traitement génotoxique (doxorubicine 0,3 μ M). Cette analyse a été réalisée à l'aide de l'outil DESeq2 avec comme filtres une p-value inférieure à 5 % et un log2Fold change supérieur à 2 ou inférieur à -2. Avec ces critères, nous avons mis en évidence 736 gènes différentiellement exprimés entre le groupe témoin et le groupe des mutants *TP53*. Parmi ces gènes, un tiers correspond à des gènes d'ARN non-codants. 566 sont sous-exprimés chez les mutants par rapport aux contrôles alors que 170 sont sur-exprimés chez les mutants. Une analyse ontologique réalisée sur les 736 gènes différentiellement exprimés révèle une multitude de voies biologiques différentes parmi lesquelles la neurogenèse, la migration cellulaire, l'angiogenèse ou l'apoptose (**Figure 43**).

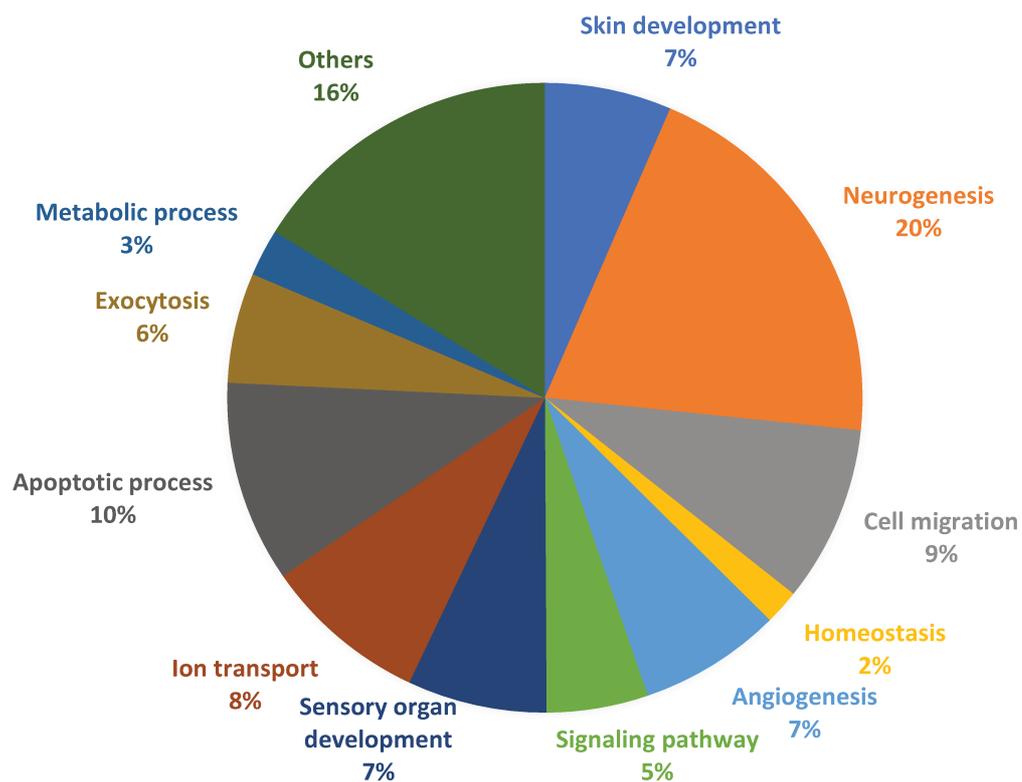


Figure 43 : Analyse d'Ontologie réalisée sur les gènes différentiellement exprimés en présence de la doxorubicine entre les lignées mutées pour *TP53* et les lignées témoins

- **Création d'un modèle prédictif (analyse OPLS-DA)**

Lorsque nous réalisons une analyse comparative supervisée avec l'outil OPLS-DA sur les échantillons traités à la doxorubicine, la différence observée entre les groupes de témoins et de mutants pour *TP53* est encore augmentée par rapport à l'état basal avec une p-value de l'ordre de 10^{-9} (Figure 44). Ces résultats très encourageants semblent indiquer qu'il serait possible d'utiliser les données de RNA-seq pour identifier les porteurs de mutations délétères de *TP53*. Ces résultats restent à confirmer sur une plus grande série d'échantillons.

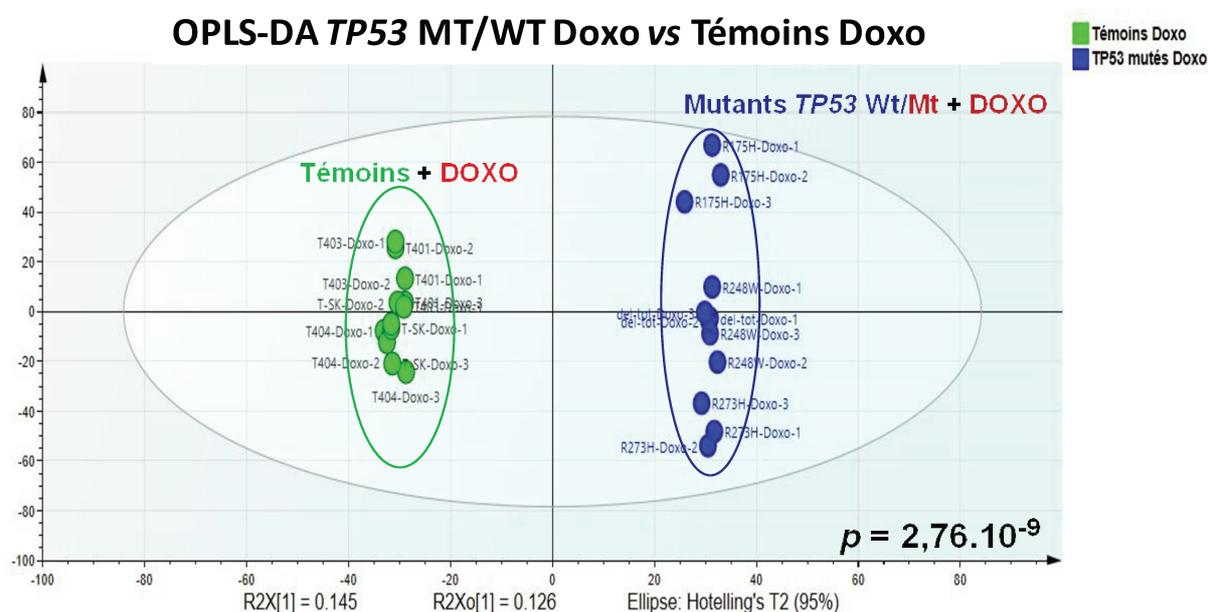


Figure 44 : Analyse OPLS-DA des données de RNA-Seq obtenues pour les lignées mutées pour le gène *TP53* et les lignées témoins en présence de la doxorubicine

Séparation des échantillons témoins (en vert) et des mutés pour le gène *TP53* (en bleu) avec traitement génotoxique (doxorubicine $0,3 \mu\text{M}$ pendant 8h).

B- Comparasion des lignées BRCA Wt/Mt et des lignées témoins

Nous avons dans un second temps analysé le transcriptome des cellules avec une mutation délétère d'un des gènes *BRCA* avec celui des cellules sauvages. Pour cette deuxième comparaison, seule la condition avec traitement à la bléomycine a pu être analysée en RNA-Seq. Une analyse en composante principale ainsi qu'une analyse de clustering hiérarchique ont été réalisées. L'analyse en composante principale ne met pas en évidence de séparation évidente des différents génotypes (données non présentées). Le clustering hiérarchique met en évidence que le premier critère de différenciation des échantillons est principalement la lignée cellulaire, puis le traitement. Il n'apparaît pas de différenciation claire des groupes témoins ou mutants dans ces analyses non supervisées (**Figure 45**).

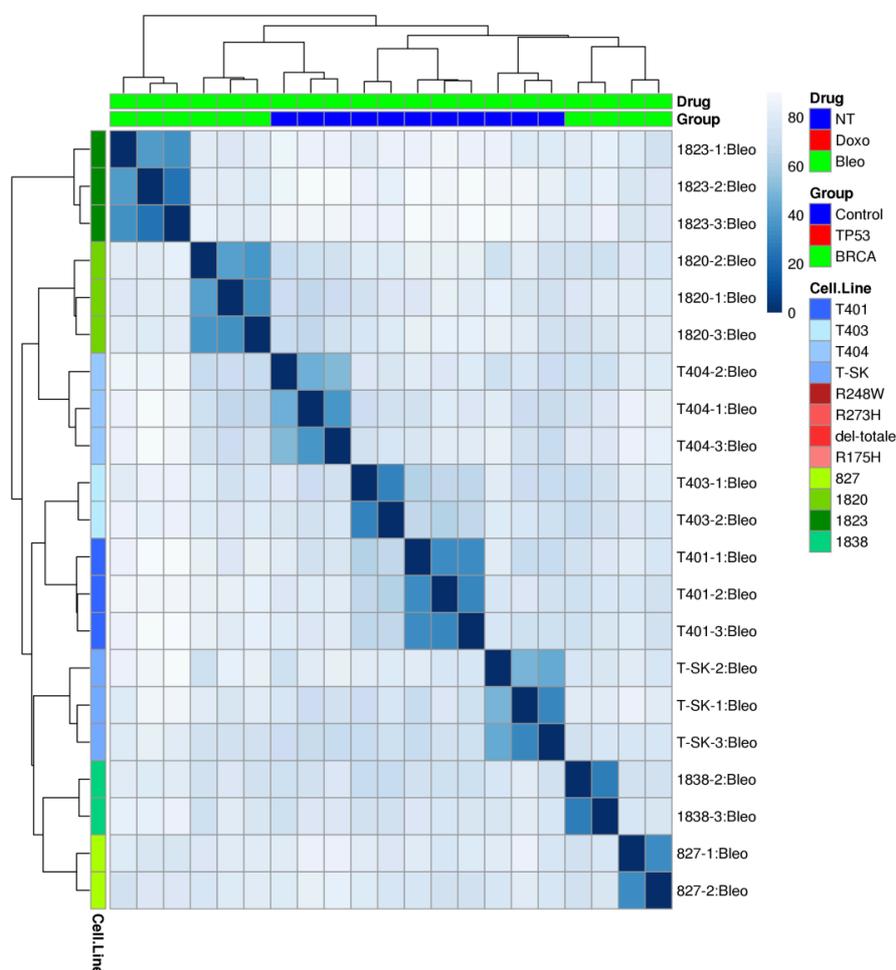


Figure 45 : Analyse de clustering hiérarchique (non supervisée) des données de RNA-Seq obtenues pour toutes les lignées témoins et les lignées mutées BRCA après traitement à la bléomycine

- Analyse d'expression différentielle

Nous avons réalisé une analyse d'expression différentielle entre les lignées sauvages ou mutantes pour l'un des gènes *BRCA* en présence d'un traitement génotoxique (bléomycine 7 μ M). Cette analyse a été réalisée à l'aide de l'outil DESeq2 avec comme filtres une p-value inférieure à 5 % et un log2Fold change supérieur à 2 ou inférieur à -2. Avec ces critères, nous avons mis en évidence 413 gènes différentiellement exprimés entre le groupe témoin et le groupe des mutants BRCA. Parmi ces gènes, 217 sont sous-exprimés chez les mutants par rapport aux contrôles alors que 196 sont sur-exprimés chez les mutants. Une analyse ontologique réalisée sur les 413 gènes différentiellement exprimés révèle que ces gènes sont reliés à une multitude de voies biologiques différentes parmi lesquelles l'angiogenèse, la transduction du signal, l'organisation des membranes ou la neurogène (Figure 46).

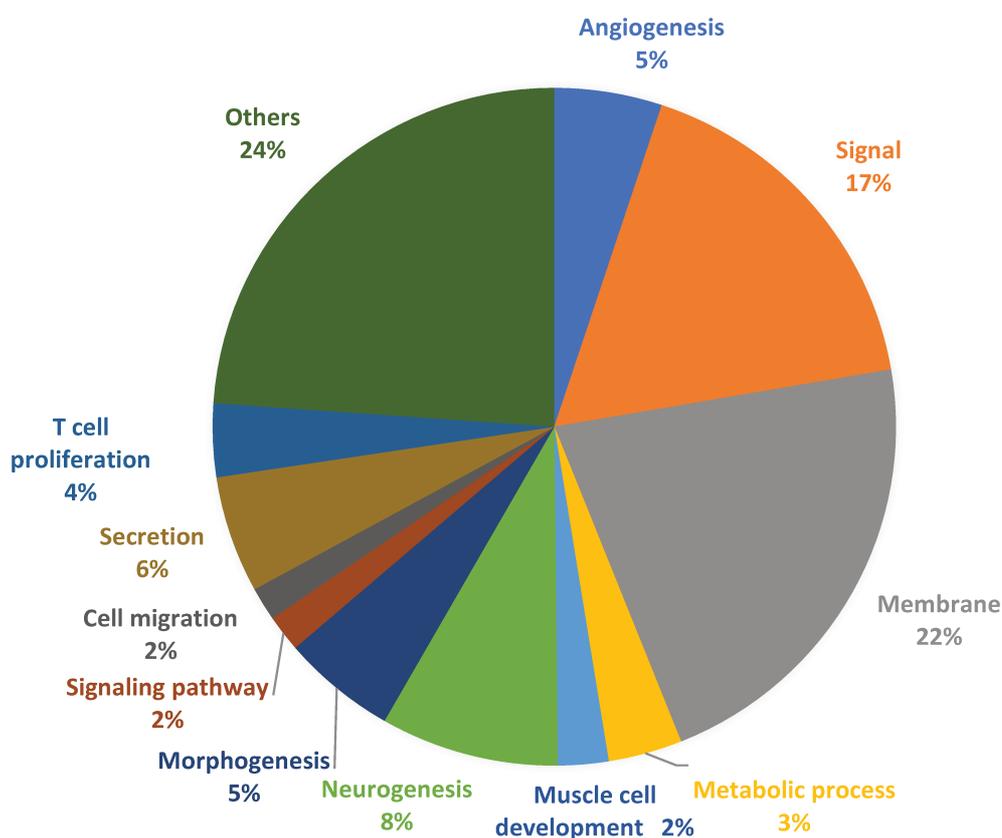


Figure 46 : Analyse d'Ontologie réalisée sur les gènes différentiellement exprimés en présence de la bléomycine entre les lignées mutées pour l'un des gènes *BRCA* et les lignées témoins

- Création d'un modèle prédictif (analyse OPLS-DA)

Lorsque nous réalisons une analyse comparative supervisée avec l'outil OPLS-DA sur les échantillons traités à la bléomycine, les groupes de témoins et de mutants pour *BRCA* se séparent avec une p-value de l'ordre de 10^{-7} (Figure 47). Ces résultats encourageants restent à confirmer sur une plus grande série d'échantillons.

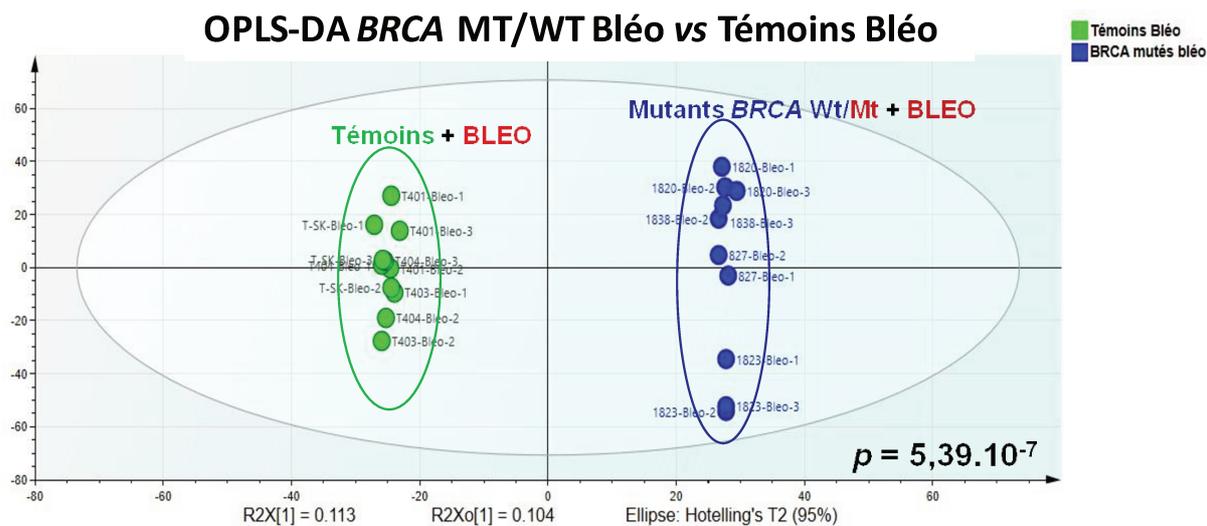


Figure 47 : Analyse OPLS-DA des données de RNA-Seq obtenues en présence de la bléomycine pour les lignées mutées dans l'un des gènes *BRCA* et les lignées témoins

Séparation des échantillons témoins (en vert) et des mutés pour l'un des gènes *BRCA* (en bleu) avec traitement génotoxique (bléomycine 7 μ M pendant 8h).

4- Sélection de biomarqueurs potentiels et développement de tests fonctionnels applicables au gène TP53

a- Développement d'un test fonctionnel TP53 à l'état basal

Notre deuxième objectif était de dériver à partir des données omiques obtenues un test fonctionnel simplifié qui soit adaptable à la routine diagnostique. Nous nous sommes focalisés sur les données TP53. Le test fonctionnel dédié à TP53 décrit dans la première partie de ma thèse s'est avéré être un outil efficace pour déterminer l'impact des variations détectées dans TP53 et pour détecter un éventuel effet dominant négatif de certaines mutations. La première limite de ce test réside dans la nécessité de disposer pour chaque patient d'une lignée lymphoblastoïde. L'établissement d'une lignée représente un coût et nécessite plusieurs mois de culture. Compte-tenu de l'impact sur la prise en charge du patient que représente aujourd'hui l'identification d'une mutation délétère de TP53, ce délai devient difficilement compatible avec la nécessité d'interpréter rapidement tous les VSI qui seront identifiées dans ce gène. De plus, suite à l'intégration récente du gène TP53 aux panels de gènes dédiés à la prédisposition génétique au cancer du sein et de l'ovaire, ce nombre de VSI va augmenter de façon exponentielle soulignant le besoin de tests encore plus rapides.

Compte-tenu des résultats d'expression différentielle obtenus à l'état basal pour les lignées mutées pour TP53, nous avons dans un premier temps exploré la possibilité d'extraire des biomarqueurs différentiellement exprimés entre les cellules sauvages et les cellules mutées à l'état hétérozygote pour les inclure dans un test simplifié. En effet, les résultats obtenus en RNA-Seq laissent supposer qu'il existe un endophénotype détectable des cellules mutées à l'état hétérozygote même sans induction de traitement génotoxique. Sur la base de l'analyse différentielle réalisée avec DeSEQ2, nous avons extrait **6 gènes** différentiellement exprimés entre le groupe des témoins et le groupe des mutants TP53 : PTPRG, RNF180, TLN2, MGST1, SPATA18 et PDE3B dont la fonction est précisée dans le **tableau 7** et dont le nombre de reads normalisés obtenu est illustré dans la **Figure Annexe 1**.

Gène	Fonction de la protéine codée par le gène	NM	MIM
PTPRG (Protein Tyrosine Phosphatase, Receptor Type G)	Tyrosine phosphatase régule une variété de processus cellulaires directement liés au cancer, tels que la prolifération, la différenciation, l'apoptose, la migration et l'invasion (Liu et al., 2016)	NM_002841	176886
TLN2 (Talin 2)	Protéine du cytosquelette impliquée dans l'assemblage des filaments d'actine et dans la propagation et la migration de divers types de cellules (Qi et al., 2016)	NM_015059	607349
RNF180 (Ring Finger Protein 180)	Ubiquitine ligase E3 impliquée dans la prolifération et la différenciation et dans l'organisation des membranes cellulaires (Nakamura, 2011)	NM_0001113561	616015
MGST1 (Microsomal Glutathione S-Transferase 1)	Enzyme impliquée dans la défense cellulaire contre les stress oxydatifs, dans l'hématopoïèse et dans le développement embryonnaire (Bräutigam et al., 2018)	NM_001260511	138330
SPATA18 (Spermatogenesis Associated 18)	Protéine impliquée dans la spermatogenèse, dans le contrôle de qualité mitochondrial en induisant l'élimination des protéines mitochondriales oxydées (Bornstein et al., 2011) (Kitamura et al., 2011)	NM_145263	612814
PDE3B (Phosphodiesterase 3)	Phosphodiesterase impliquée dans de nombreux processus physiologiques et dans le métabolisme des tissus adipeux et dans la modulation de la réponse inflammatoire (Ahmad et al., 2016)	NM_000922	602047

Tableau 7 : Fonctions des 6 gènes sélectionnés sur la base de l'analyse d'expression différentielle DeSEQ2 des données de RNA-Seq et analysés en RT-QMPSF

Nous avons ensuite intégré ces biomarqueurs dans une réaction ciblée d'expression génique de RT-QMPSF. Ces gènes étaient comparés à un gène témoin TBP.

Après mise au point des conditions de RT-QMPSF, nous avons confirmé les résultats de RNA-Seq obtenus pour ces 6 gènes en comparant les 4 lignées témoins avec les 4 lignées mutées *TP53* (**Figure 20**). Les résultats montrent comme attendu que les niveaux d'expression des gènes *PTPRG*, *RNF180*, *TLN2*, *MGST1* et *SPATA18* sont plus élevés chez les témoins que chez les mutants et que celui du gène *PDE3B* est plus élevé chez les mutants. Après la vérification des résultats de RNA-Seq par cette analyse ciblée, nous avons voulu valider notre essai en analysant une collection de 30 lignées lymphoblastoïdes établies à partir de patients LFS porteurs de mutations constitutionnelles de *TP53*. Cette collection exceptionnelle comprend différents types de mutations de *TP53*, parmi lesquelles de nombreuses mutations faux-sens avec des effets dominants négatifs connus, ainsi que des mutations frameshift, non-sens, des mutations d'épissage et des réarrangements géniques.

Après analyse des 30 lignées avec notre réaction de RT-QMPSF dédiée à l'état basal, il apparaît que les niveaux d'expression de ces gènes n'étaient pas les mêmes chez les patients portant les autres types de mutations *TP53*. En effet, la surexpression des gènes *PTPRG*, *RNF180*, *TLN2*, *MGST1* et *SPATA18* observée chez les 4 mutants analysés en RNA-seq n'est pas observée chez tous les autres mutants et il en est de même pour l'expression gène *PDE3B* (**Figure 49**). Les différences d'expression observées pour ces 6 gènes entre les témoins et les mutants analysés en RNA-Seq sont donc le reflet de la très grande variabilité inter-individuelle de l'expression génique. Le faible effectif analysé en RNA-Seq ne nous permet pas de nous affranchir de ce bruit de fond et devra être par la suite augmenté. Ainsi, ces gènes choisis ne peuvent pas être utilisés comme biomarqueurs discriminants.

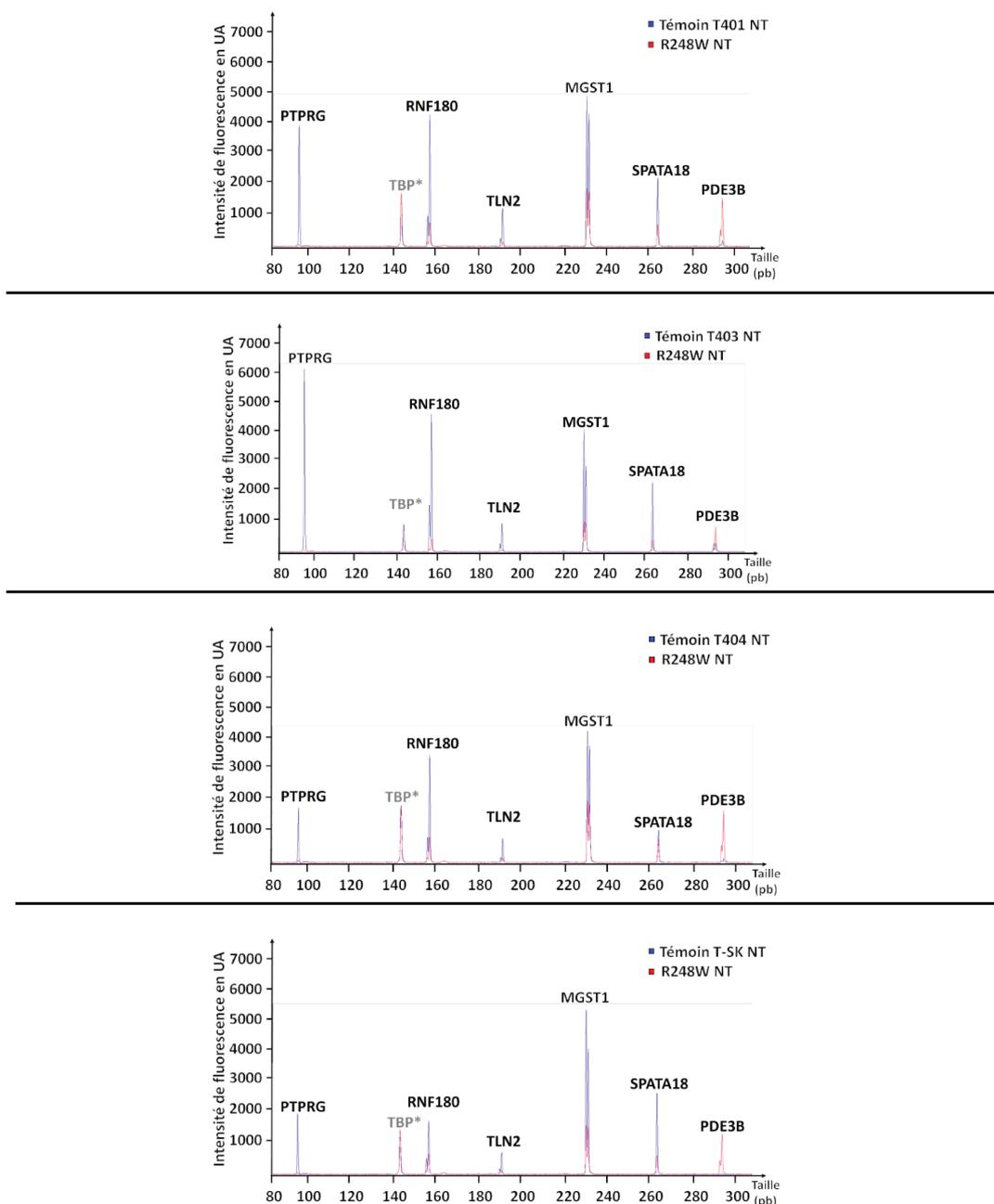


Figure 48 : Analyse par RT-QMSPF des niveaux relatifs d'expression à l'état basal des gènes sélectionnés par RNA-Seq dans la lignée immortalisée R248W du gène *TP53*

Le taux relatif d'ARN messager de 6 gènes différemment exprimés entre les cellules sauvages (en bleu) ou les cellules portant la mutation **R248W** à l'état hétérozygote de *TP53* (rouge) a été évalué par RT-QMSPF après superposition des profils de fluorescence et ajustement sur la hauteur du gène contrôle (TBP). Les différents gènes sont indiqués. L'axe vertical indique l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paires de bases (pb). NT : Non traité.

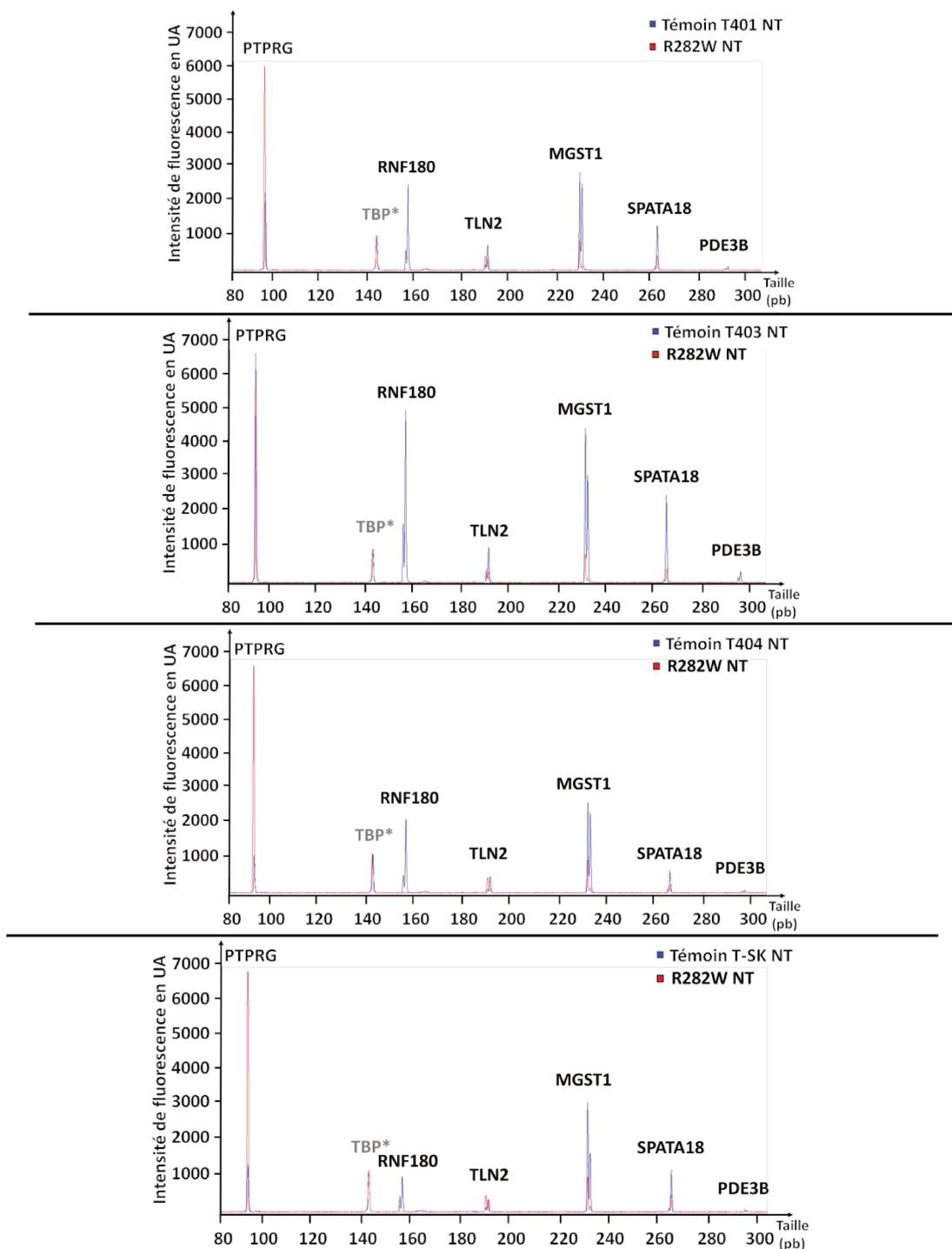


Figure 49 : Analyse par RT-QMSPF des niveaux relatifs d'expression à l'état basal des gènes sélectionnés par RNA-Seq dans la lignée immortalisée R282W du gène *TP53*

Le taux relatif d'ARN messager de 6 gènes différentiellement exprimés entre les cellules sauvages (en bleu) ou les cellules portant la mutation **R282W** à l'état hétérozygote de *TP53* (rouge) a été évalué par RT-QMSPF après superposition des profils de fluorescence et ajustement sur la hauteur du gène contrôle (TBP). Les différents gènes sont indiqués. L'axe vertical indique l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paires de bases (pb). NT : Non traité.

b- Développement d'un nouvel essai fonctionnel TP53 applicable sur prélèvement sanguin

Le développement d'un test fonctionnel *TP53* simplifié et réalisé sans traitement génotoxique n'étant pas possible en l'état actuel du projet, nous avons décidé d'explorer la possibilité de s'affranchir de la disponibilité d'une lignée lymphoblastoïde pour réaliser le test fonctionnel directement sur prélèvement sanguin ce qui rendrait les délais du test fonctionnel plus compatible avec la routine diagnostique.

i. Développement d'un nouvel essai fonctionnel TP53 basé sur la RT-MLPA

Nous avons dans un premier temps exploité les données de RNA-Seq obtenues avec le traitement génotoxique pour faire évoluer la première version du test fonctionnel *TP53* décrite dans la première partie du manuscrit. Nous avons choisi d'ajouter plus de gènes cibles de *TP53* impliqués dans différentes voies biologiques. Pour cela, nous avons sélectionné des gènes dont l'expression est induite par le traitement à la doxorubicine chez les témoins et qui sont peu ou pas induits lorsque *TP53* est muté à l'état hétérozygote. Les 10 gènes retenus sont KIAA0284, PODXL, RRAD, GLS2, CABYR, TP53I3, EPS8L2, SULF2, SESN1, FHL2 dont les fonctions sont résumées dans le **Tableau 8** et le nombre de reads obtenus illustrés dans la **Figure Annexe 2**. Nous avons également sélectionné sur la base des comptages de reads des gènes témoins supplémentaires pour normaliser de façon plus robuste nos analyses. Les trois gènes témoins dont l'expression semble constante entre toutes les conditions sont TBP, RIC8B et MPP5 (**Figure Annexe 3**). Pour vérifier le niveau d'expression du gène *TP53* dans les lignées analysées nous avons choisi d'intégrer deux marqueurs des jonctions des exons 4-5 et 6-7 de *TP53*. Le ratio obtenu entre la hauteur du pic pour la jonction *TP53* analysée et la hauteur des pics témoins pourrait être comparé entre un individu témoin et un patient avec variant de *TP53*.

Gène	Fonction de la protéine	NM	MIM
<i>PODXL</i> (<i>Podocalyxin-like 1</i>)	Impliquée dans l'adhésion, la morphologie et la migration cellulaire (Zhang <i>et al.</i> , 2018)	NM_001018111	602632
<i>KIAA0284</i>	Joue un rôle dans l'organisation des microtubules (Welburn & Cheeseman, 2012)	NM_015005	-
<i>RRAD</i> (<i>Ras related glycolysis inhibitor and calcium channel regulator</i>)	Rôle important dans la régulation de la glycolyse aérobie, dans l'invasion cellulaire et dans les métastases (Shang <i>et al.</i> , 2016)	NM_004165	179503
<i>EPS8L2</i> (<i>Epidermal Growth Factor Receptor Pathway Substrate 8-Related Protein 2</i>)	Impliquée dans le remodelage cytosquelettique de l'actine au niveau des membranes cellulaires (Offenhäuser <i>et al.</i> , 2004)	NM_022772	614988
<i>GLS2</i> (<i>Glutaminase 2</i>)	Rôle important dans la régulation du catabolisme de la glutamine, favorise la respiration mitochondriale et augmente la fonction anti-oxydante de la cellule (Xiang <i>et al.</i> , 2013)	NM_013267	606365
<i>SULF2</i> (<i>Sulfatase 2</i>)	Impliquée dans le métabolisme des lipoprotéines, dans l'angiogenèse et dans la signalisation cellulaire (Dallinga & Dallinga-Thie, 2016)	NM_001161841	610013
<i>CABYR</i> (<i>Calcium Binding Tyrosine Phosphorylation Regulated</i>)	Impliquée dans la formation de la gaine fibreuse essentielle à la fertilité masculine (Young <i>et al.</i> , 2016)	NM_012189	612135
<i>TP53I3</i> (<i>Tumor Protein p53 Inducible protein 3</i>)	Impliquée dans la réponse cellulaire aux stress oxydatifs et à l'irradiation et dans l'apoptose (Quan <i>et al.</i> , 2017)	NM_004881	605171
<i>SESNI</i> (<i>Sestrin 1</i>)	Rôle dans la réponse cellulaire aux lésions de l'ADN et au stress oxydatif (Kabacik <i>et al.</i> , 2011)	NM_014454	606103
<i>FHL2</i> (<i>Four and a Half Lim Domains 2</i>)	Rôle dans l'organisation des membranes extracellulaires, la signalisation cellulaire, la régulation de la transcription et des cellules souches hématopoïétiques ainsi que leur auto-renouvellement (Hou <i>et al.</i> , 2015)	NM_201555	602633

Tableau 8 : Fonctions des 10 gènes sélectionnés sur la base de l'analyse d'expression différentielle DESeq2 des données de RNA-Seq et analysés en RT-MLPA

Nous avons choisi d'intégrer ces biomarqueurs dans un test de RT-MLPA (Reverse Transcriptase-Multiplex Ligation-dependent Probe Amplification). Cette technique utilise le principe de la MLPA initialement développé pour détecter les variations du nombre de copies des gènes et combine cette approche avec une phase de reverse-transcription. Après production de l'ADN complémentaire, 2 sondes par cible sont hybridées de façon juxtaposée. Si la cible est présente, une étape de ligation permettra de liguer les deux sondes ensemble. L'ensemble des sondes liguées correspondant à chaque cible est ensuite amplifié à l'aide d'un couple de primers universels complémentaires des séquences additionnelles ajoutées aux sondes et dont le primer sens est marqué par un fluorochrome (**Figure 50**). Les réactions de RT-MLPA, sont ensuite analysées sur un séquenceur automatique et les différents amplicons fluorescents sont séparés par taille. L'ajout d'une séquence « spacer » aux sondes utilisées permet de créer artificiellement une différence de taille entre les produits de ligation. Cette différence de taille maîtrisée permet d'analyser un nombre important de cibles en une seule réaction. Cette méthode rapide permet d'étudier l'expression relative d'une vingtaine de transcrits d'intérêt en une seule réaction (Kim et al., 2015b). La RT-MLPA présente l'avantage de pouvoir être appliquée de façon robuste même lorsque l'ARN n'est pas de très bonne qualité comme lorsque l'ARN est extrait à partir de tissus fixés et inclus en paraffine (Bobée et al., 2017). En effet, la première étape est une étape d'hybridation et non d'amplification qui peut être plus sensible à la qualité de la matrice.

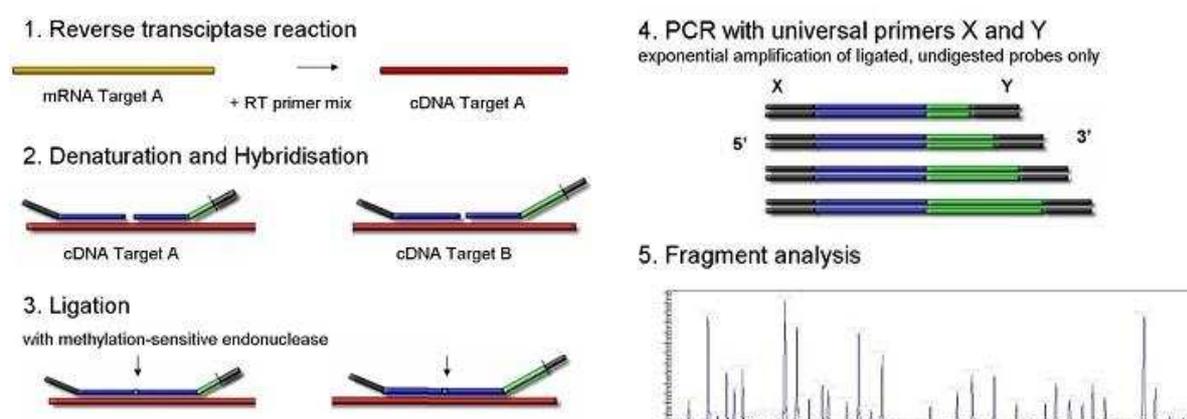


Figure 50 : Principe de la technique de RT-MLPA (D'après <https://www.mlpa.com>)

Après la synthèse de l'ADNc par une réaction de reverse transcription, la RT-MLPA se poursuit en commençant par l'hybridation des sondes à leur ADNc cible. Les sondes pour chaque gène cible sont choisies pour inclure une jonction exon-exon du gène dans leur

séquence cible. Elles contiennent également à leurs extrémités deux queues universelles permettant d'effectuer par la suite l'étape d'amplification par PCR. Après amplification avec une seule paire d'amorces fluorescentes universelles, les échantillons contenant des fragments d'ADN mélangés de différentes tailles sont quantifiés sur un séquenceur capillaire.

Après mise au point des conditions, nous avons appliqué ce test ciblé de RT-MLPA sur nos 4 lignées témoins ainsi que sur les 34 lignées issues de patients LFS porteurs d'une mutation à l'état hétérozygote de TP53. Après superposition des profils de RT-MLPA obtenus pour un individu à partir des lymphocytes traités et non traités à la doxorubicine, nous avons ajusté les hauteurs de profil sur nos 3 gènes contrôles. Nous avons ensuite calculé le score de fonctionnalité de p53, correspondant à la moyenne du taux d'induction des 10 gènes étudiés. Les résultats montrent la forte induction des 10 gènes cibles sélectionnés dans les lymphocytes de 4 sujets contrôles de génotype TP53 sauvage (**Figure 51**) avec un score de fonctionnalité égal à $11,7 \pm 3,5$. Les 10 patients porteurs de mutations perte de fonction (« nulles ») présentent dans ce test une induction intermédiaire avec un score de fonctionnalité de $5,5 \pm 1,4$. Les 9 patients porteurs de mutations faux-sens connues pour avoir un effet dominant négatif présentent quant à eux une forte réduction du taux d'induction de ces gènes avec un score de fonctionnalité de $3,8 \pm 1,4$ (**Figure 52**). Pour les 15 mutations faux-sens sans effet dominant négatif connu, le score de fonctionnalité moyen obtenu est de $5,3 \pm 2$. Les résultats obtenus sont résumés dans le **Tableau 9**. Nous avons ainsi pu valider l'efficacité de notre nouvel essai fonctionnel.

De façon parallèle, avec l'aide de Marion Rolain, ingénieur d'étude dans notre laboratoire, nous avons développé un essai fonctionnel basé sur la RT-QMPSF et incluant toutes les nouvelles cibles (**Figure 53**). Cet essai a été utilisé pour analyser en parallèle toutes les lignées testées en RT-MLPA. Les résultats obtenus avec les deux techniques sont concordants et nous pourrions donc bénéficier de deux tests évaluant les mêmes cibles mais basés sur des techniques différentes (**Tableau 9**).

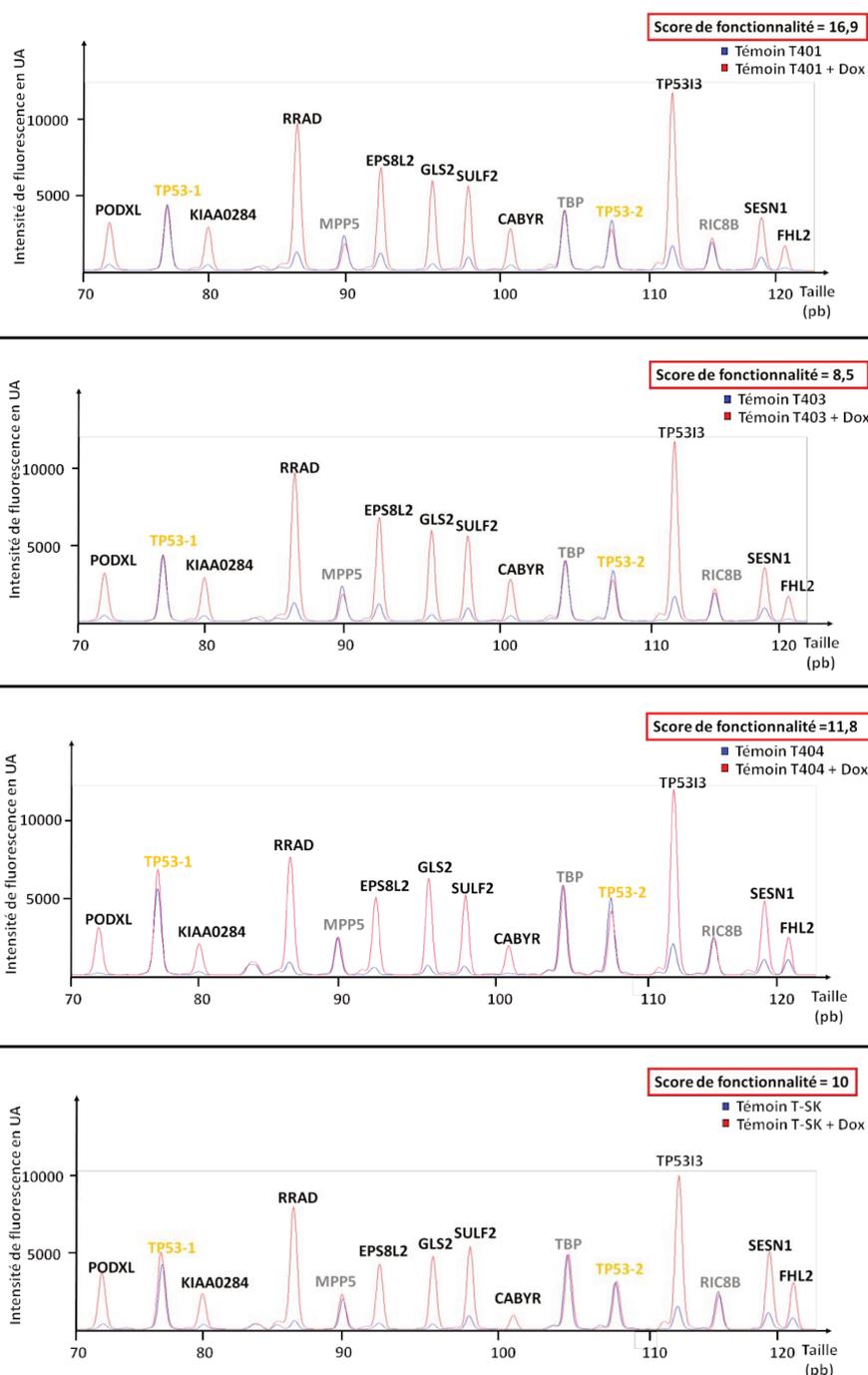


Figure 51 : Évaluation du score de fonctionnalité de p53 grâce à l'essai fonctionnel de RT-MLPA

Les profils de RT-MLPA des lymphocytes immortalisés traités (profil rouge) et non traités (profil bleu) de sujets contrôles ont été superposés. Les différents gènes cibles (PODXL1, KIAA0284, RRAD, EPS8L2, GLS2, SULF2, CABYR, TP53I3, SESN1, FHL2), gènes contrôles (MPP5, TBP, RIC8B) et les amplicons du gène TP53 (en jaune) sont indiqués. L'axe vertical montre l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paire de bases (pb) ; Dox : Doxorubicine

A) Score du sujet témoin T401 ; B) Score du sujet témoin T403 ; C) Score du sujet témoin T404 ; D) Score du sujet témoin T-SK.

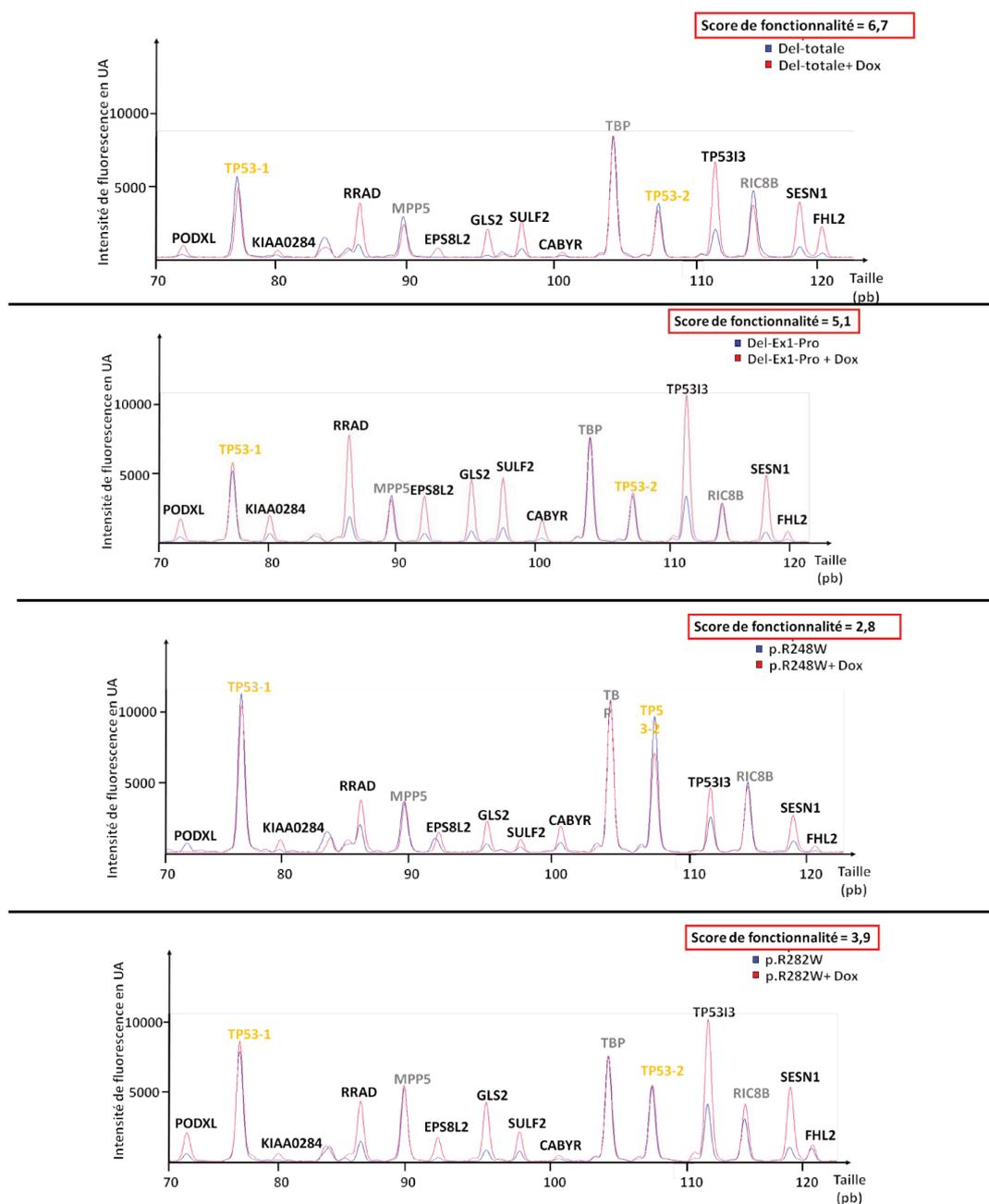


Figure 52 : Evaluation du score de fonctionnalité de p53 grâce à l'essai fonctionnel de RT-MLPA

Les profils de RT-MLPA des lymphocytes immortalisés traités (profil rouge) et non traités (profil bleu) de sujets mutants pour TP53 ont été superposés. Les différents gènes cibles (PODXL1, KIAA0284, RRAD, EPS8L2, GLS2, SULF2, CABYR, TP5313, SESN1, FHL2), gènes contrôles (MPP5, TBP, RIC8B) et les amplicons du gène TP53 (en jaune) sont indiqués. L'axe vertical montre l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paire de bases (pb). Dox : Doxorubicine

A et B : Scores de deux mutations nulles ; C et D : Scores de deux mutations faux-sens avec un effet dominant négatif connu.

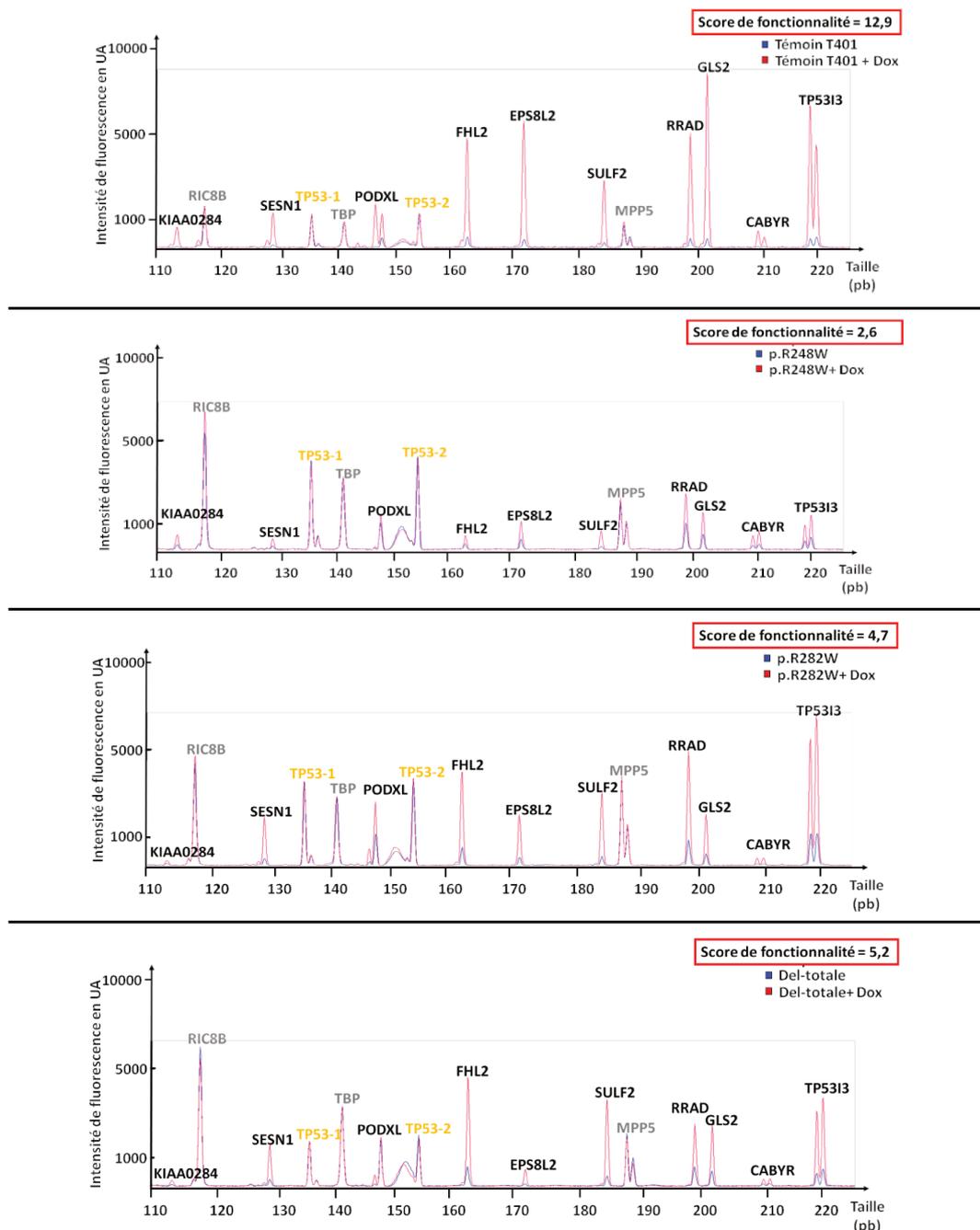


Figure 53 : Evaluation du score de fonctionnalité de p53 grâce à l'essai fonctionnel de RT-QMPSF

Les profils de RT-QMPSF des lymphocytes immortalisés traités (profil rouge) et non traités (profil bleu) de sujets contrôles et mutants pour TP53 ont été superposés. Les différents gènes cibles (PODXL1, KIAA0284, RRAD, EPS8L2, GLS2, SULF2, CABYR, TP5313, SESN1, FHL2), gènes contrôles (MPP5, TBP, RIC8B) et les amplicons du gène TP53 (en jaune) sont indiqués. L'axe vertical montre l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paire de bases (pb) ; Dox : Doxorubicine

A) Score du sujet témoin T401 ; B et C) Scores de deux mutations faux-sens avec un effet dominant négatif connu ; D) Score d'une mutation nulle.

N° Patient	Génotype	Protéine mutante	Exon	Domaine	Score RT-QMPSF (6 gènes)	Score RT-QMPSF (10 gènes)	Score RT-MLPA (10 gènes)	Ratio TP53 - 1	Ratio TP53 - 2
TEMOINS									
T401	TP53 Wt	NA	NA	NA	13,1	12,6	16,6	0,37	0,39
T403	TP53 Wt	NA	NA	NA	11	8,5	8,5	0,33	0,36
T404	TP53 Wt	NA	NA	NA	14,6	15,3	11,8	0,43	0,48
T-SK	TP53 Wt	NA	NA	NA	11,6	13,6	10	0,27	0,29
MUTATIONS FAUX SENS AVEC EFFET DOMINANT NEGATIF CONNU									
12459.003	c.524G>A	p.R175H	5	DBD	4,1	5,7	4,4	0,43	0,45
19963.001	c.742C>T	p.R248W	7	DBD	3,2	2,6	2,8	0,38	0,39
26683.001	c.818G>A	p.R273H	8	DBD	1,9	3,3	5,2	0,33	0,33
14624.008	c.844C>T	p.R282W	8	DBD	4,6	4,7	3,9	0,32	0,33
23208.001	c.733G>A	p.G245S	7	DBD	2,4	2,4	2,7	0,32	0,37
28850.001	c.638G>A	p.R213Q	6	DBD	4,1	5,3	6,7	0,33	0,44
4949.006	c.535C>T	p.H179Y	5	DBD	3,8	3	3,9	0,37	0,40
30460.001	c.736A>G	p.M246V	7	DBD	2,7	2,7	3,2	0,36	0,43
31214.001	c.743G>A	p.R248Q	7	DBD	2,2	2,1	2,2	0,30	0,35
MUTATIONS PERTE DE FONCTION									
5415.013	NA	délétion totale	NA	DBD	7,2	5,2	6,7	0,17	0,19
25771.001	c.632_641del	p.T211Ifs*33	6	DBD	5,7	4,9	5,1	0,21	0,23
23165.001	c.820del	p.V274Ffs*71	8	DBD	7	7,1	5,1	0,20	0,18
7656.006	c.(?_202)_(29+1-28-1)del	p.(del-prom-ex1)	NA	/	5,4	5,8	5,1	0,23	0,22
21057.002	c.216dup	p.V73Rfs*76	4	TAD	6,5	6	5,7	0,19	0,19
26614.001	c.323_329dup	p.L111Ffs*40	7	TAD	6,9	4,4	4,4	0,25	0,23
5534.010	c.673-2A>G	p.?	i6	DBD	5,5	5,5	5,2	0,21	0,22
28636.001	c.491_494del	p.K164Sfs*5	5	DBD	8,6	8,1	6,9	0,25	0,24
27950.001	c.455del	p.P152Rfs*18	5	DBD	4,7	3,2	3	0,24	0,24
20651.002	c.690del	p.T231Pfs*16	7	DBD	9,4	8	8	0,27	0,28
MUTATIONS FAUX SENS SANS EFFET DOMINANT NEGATIF CONNU									
8933.004	c.473G>A	p.R158H	5	DBD	5,1	5,6	6	0,33	0,37
26709.001	c.1010G>A	p.R337H	10	TET	5,8	5,3	5,3	0,34	0,38
27053.001	c.374G>A	p.T125M	4	DBD	5,1	7,6	9,2	0,37	0,36
29014.001	c.31G>C	p.E11Q	2	TAD	3	4,5	5,8	0,43	0,48
14197.001	c.910A>G	p.T304A	8	NLS	8,3	7,8	7,9	0,41	0,22
26970.009	c.761T>C	p.I254T	7	DBD	2,8	2,3	2,5	0,29	0,29
12097.007	c.904G>A	p.G302R	8	NLS	10	9,1	8,2	0,24	0,26
21810.003	c.323G>A	p.G108D	4	DBD	4,2	3,8	3,9	0,36	0,35
29722.001	c.685T>C	p.C229R	7	DBD	4,3	5,3	6	0,38	0,38
20443.001	c.646G>A	p.V216M	6	DBD	4,3	3,6	2,9	0,34	0,45
26371.001	c.577C>G	p.H193D	6	DBD	3,7	4	4,4	0,37	0,43
24939.001	c.578A>C	p.H193P	6	DBD	3,7	3,1	5,2	0,36	0,41
202.003	c.523C>G	p.R175G	5	DBD	3,2	3,2	2,9	0,30	0,35
22181.001	c.869G>A	p.R290H	8	DBD	10,9	7,9	7,2	0,36	0,38
21597.001	c.845G>C	p.R282P	8	DBD	4,7	3	3,2	0,37	0,39

Tableau 9 : Les scores obtenus par les techniques de RT-MLPA et RT-QMPSF pour les 4 lignées témoins et pour les 34 lignées immortalisées mutées pour le gène TP53

DBD : domaine de liaison à l'ADN ; NLS : signal de localisation nucléaire ; TAD : domaine de transactivation ; TET : domaine de tétramérisation ; Wt : Wild type ; NA : Non applicable. La quantité d'ARNm de p53 des échantillons testés a été comparée à celle des témoins à l'aide de deux amplicons (TP53 exon 4-5 et exon 6-7) et exprimée sous forme d'un ratio.

ii- Mise au point d'un test fonctionnel directement applicable sur prélèvement sanguin

Nous avons, dans une dernière étape, adapté ce test pour qu'il puisse être appliqué directement sur les lymphocytes non immortalisés du patient. Le défi de cette approche résidait dans la mise au point des conditions de prélèvement, de récupération et de mise en culture des lymphocytes. En effet, à l'état actuel du projet, la mise en évidence de l'endophénotype des cellules mutées nécessite l'application d'un stress génotoxique. Nous avons donc travaillé sur les conditions de prélèvement et de mise en culture nous permettant d'obtenir le maximum de lymphocytes viables et activés capable de répondre à un stress génotoxique. J'ai bénéficié pour cette partie du projet de l'aide de Marion Rolain. De façon générale, après prélèvement sanguin, les cellules mononucléées du sang périphérique sont isolées par centrifugation sur un gradient de ficoll. Les cellules sont ensuite mises en culture avec un milieu permettant l'activation et donc la division des lymphocytes. Une partie des cellules est ensuite traitée avec de la doxorubicine et les cellules sont ensuite récupérées pour une extraction ARN et un test RT-MLPA ou RT-QMPSF.

Nous avons comparé plusieurs tubes de prélèvements dont les tubes EDTA ou les tubes Héparinés. Pour cela, nous avons collecté du sang sur chacun de ces types de tubes et nous avons réalisé l'isolation des cellules mononucléées par séparation sur gradient de ficoll à J+1, J+2 et J+3 avec des tubes conservés à température ambiante pour simuler un envoi par courrier du prélèvement. Le nombre et la viabilité des cellules étaient ensuite évalués. Le prélèvement sur tube EDTA permettait de limiter la lyse cellulaire par rapport au tube hépariné lorsque le prélèvement n'était pas traité dans les 24 heures.

L'anneau lymphocytaire est alors mis en culture dans un milieu permettant l'activation des lymphocytes. Les cellules sont comptées et au moins 4 puits sontensemencés avec 1 million de cellules pour pouvoir réaliser la manipulation en duplicat technique. L'activation des lymphocytes est visualisable par l'apparition au microscope de colonies issues de l'activation d'une cellule. Le meilleur résultat d'activation a été obtenu en mettant les cellules en culture dans un milieu commercial «Chromosome Medium P» de chez AmpliTech (Compiègne, France). Le temps d'activation le plus favorable a été déterminé en réalisant l'essai fonctionnel et en monitorant l'intensité de la réponse p53 après différents temps de culture avant traitement allant de 0 à 5 jours. Nous avons retenu un temps d'activation de 48 heures. Si la durée du traitement n'a pas évolué comparée au traitement des lignées EBV, la dose de doxorubicine requise pour obtenir une réponse d'intensité

comparable à celle des lignées lymphoblastoïdes témoins a été multipliée par 4 (800 ng/ μ L versus 200 ng/ μ L). La quantité d'ARN, résultant de l'extraction d'un puits dans lequel ont été ensemencées un million de cellules, est beaucoup plus faible que celle obtenue avec des lignées et nous avons donc dû adapter le protocole d'extraction et utiliser le kit NucleoSpin® RNA XS (Macherey Nagel) avec un volume d'élution de 12 μ L.

Avec les conditions expérimentales retenues, nous avons pu prélever et analyser deux volontaires sains. Le score de fonctionnalité p53 obtenu pour ces deux témoins est de 11,7 et 8 en RT-MLPA (**Figure 54**) et de 10,8 et 8,1 en RT-QMPSF (**Figure 55**) à **J+2**. A partir de 3 jours post-prélèvement, nous ne sommes plus en mesure d'induire une réponse p53 au stress génotoxique (**Figure 56**). Cependant, les résultats obtenus à J+1 et J+2 illustrent la faisabilité du nouveau test fonctionnel TP53 réalisé directement sur prélèvement sanguin (**Figure 57**).

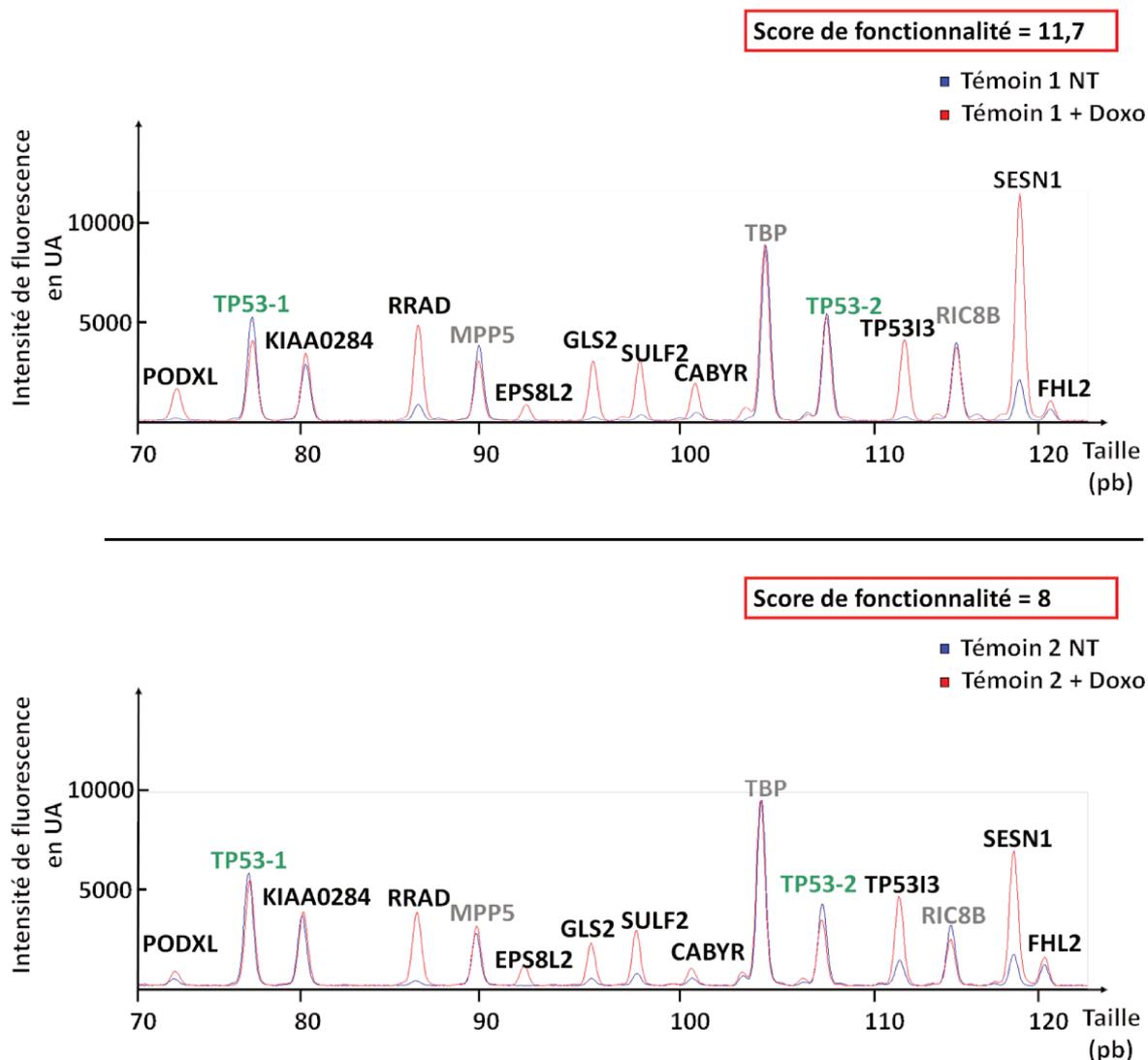


Figure 54 : Evaluation du score de fonctionnalité de p53 grâce à l’essai fonctionnel de RT-MLPA des PBMC des deux témoins pour un prélèvement à J+2

Les profils de RT-MLPA des **PBMC** traités (profil rouge) et non traités (profil bleu) de deux témoins volontaires ont été superposés. Les différents gènes cibles (PODXL1, KIAA0284, RRAD, EPS8L2, GLS2, SULF2, CABYR, TP53I3, SESN1, FHL2), gènes contrôles (MPP5, TBP, RIC8B) et les amplicons du gène TP53 (en vert) sont indiqués. L’axe vertical montre l’intensité de fluorescence en unités arbitraires (UA) et l’axe horizontal indique la taille des amplicons en paire de bases (pb).

NT : Non traité ; Doxo : Doxorubicine.

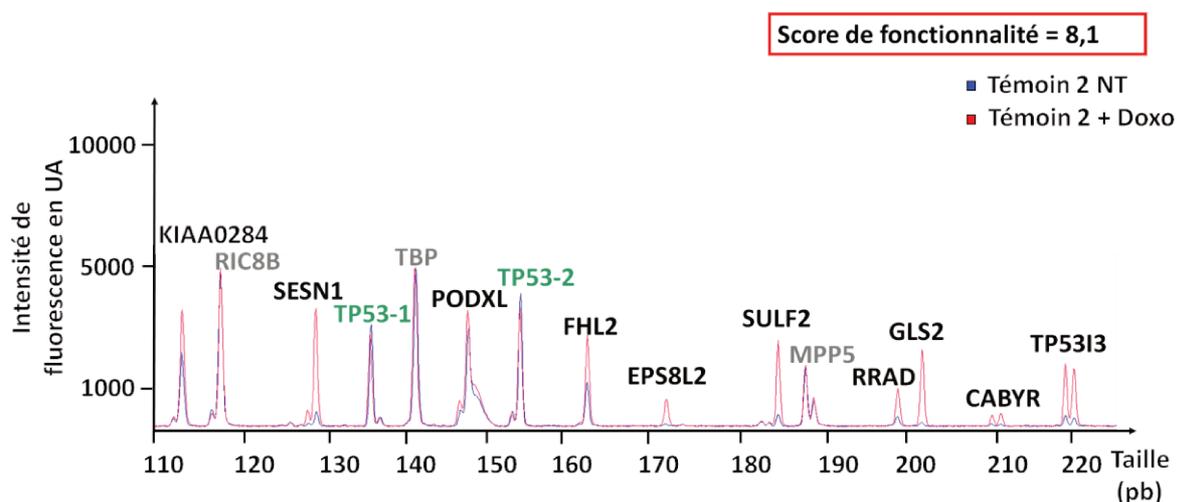
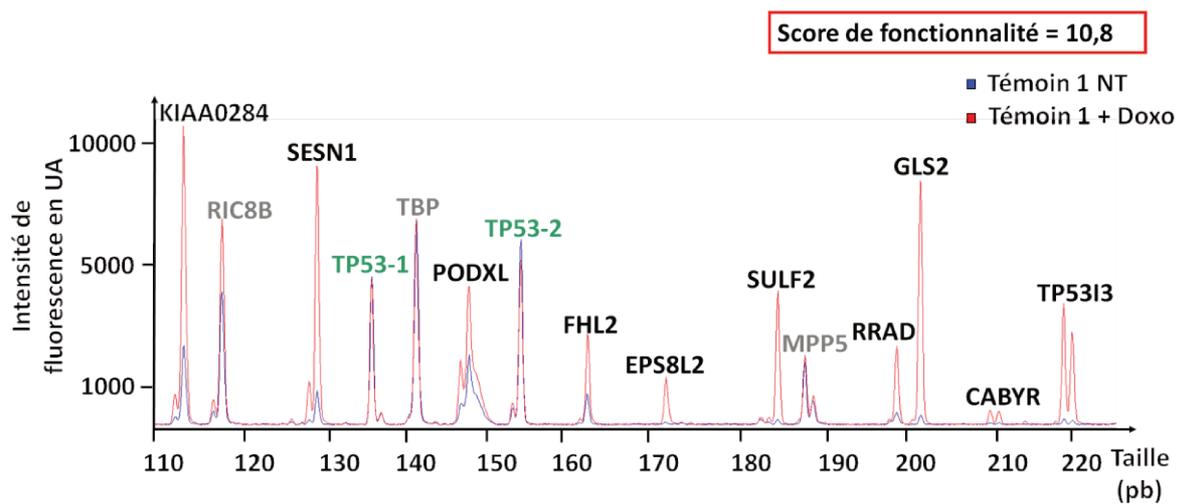


Figure 55 : Evaluation du score de fonctionnalité de p53 grâce à l'essai fonctionnel de RT-QMPSF des PBMC des deux témoins pour un prélèvement à J+2

Les profils de RT-QMPSF des PBMC traités (profil rouge) et non traités (profil bleu) de deux témoins volontaires ont été superposés. Les différents gènes cibles (PODXL1, KIAA0284, RRAD, EPS8L2, GLS2, SULF2, CABYR, TP53I3, SESN1, FHL2), gènes contrôles (MPP5, TBP, RIC8B) et les amplicons du gène TP53 (en vert) sont indiqués. L'axe vertical montre l'intensité de fluorescence en unités arbitraires (UA) et l'axe horizontal indique la taille des amplicons en paire de bases (pb).

NT : Non traité ; Doxo : Doxorubicine.

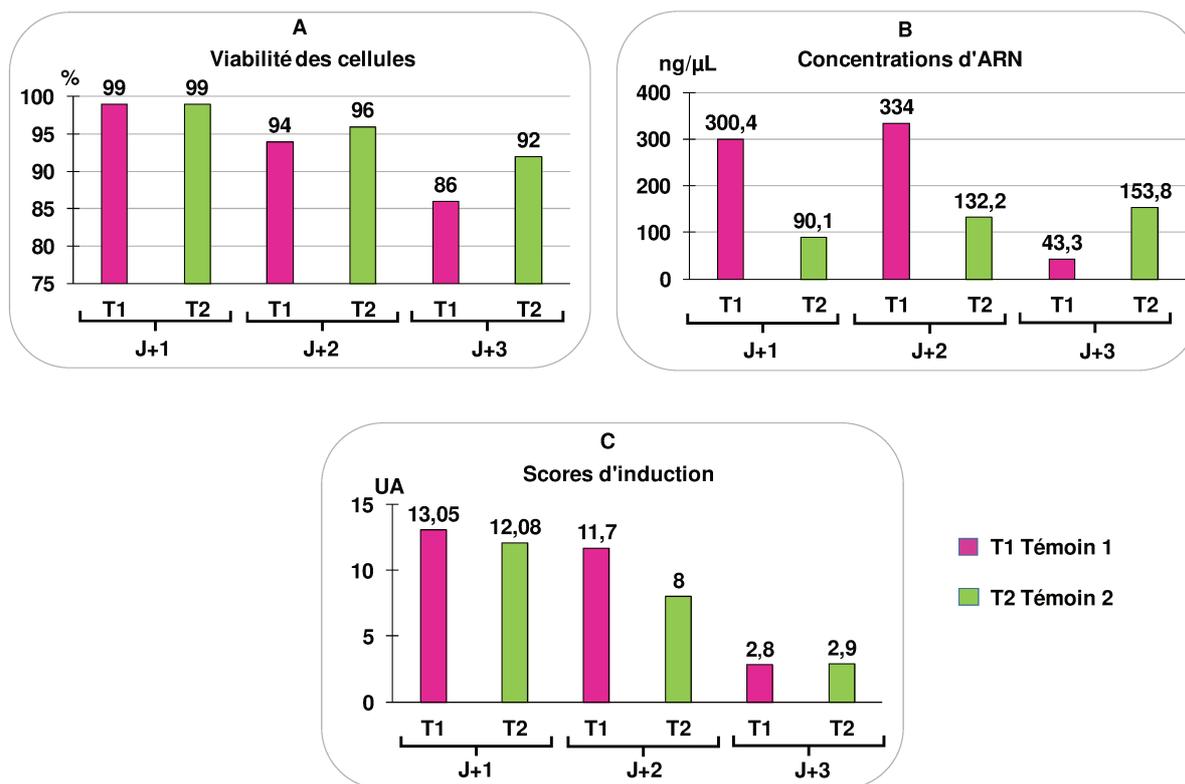


Figure 56 : Suivi des étapes du nouveau test fonctionnel *TP53* réalisé directement sur un prélèvement sanguin

(A) La viabilité des cellules (en %), (B) la concentration d'ARN (en ng/μL) et (C) les scores d'induction obtenus en RT-MLPA (en unité arbitraire, UA) après des prélèvements de PBMC effectués à J+1, J+2 et J+3 chez les deux témoins T1 et T2

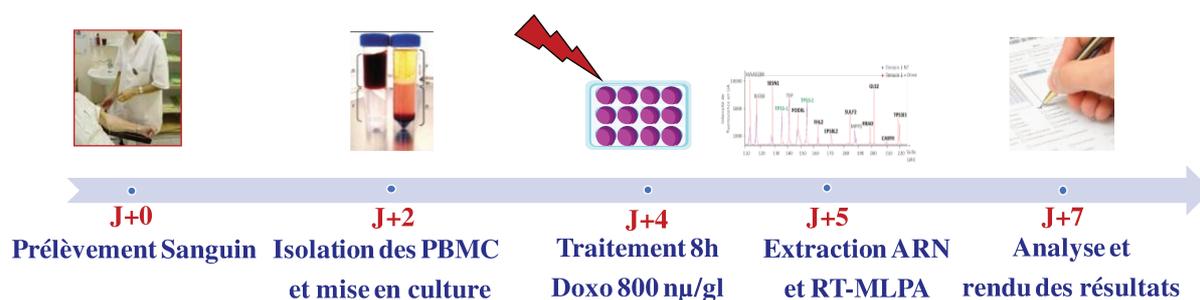


Figure 57 : Suivi des étapes du nouveau test fonctionnel *TP53* réalisé directement sur un prélèvement sanguin

IV- Discussion

Les protéines p53, BRCA1 et BRCA2 jouent un rôle central dans de nombreuses voies critiques qui convergent vers le maintien de l'intégrité génomique. L'hypothèse que nous avons émise pour ce travail est celle de l'existence d'un endophénotype des cellules non cancéreuses, mutées à l'état hétérozygote. En effet, il est très probable que la perte d'une seule copie de ces gènes puisse avoir un impact mesurable sur la fonction de la cellule. Plusieurs études ont soulevé la question de l'haploinsuffisance, provoquée par des mutations hétérozygotes des gènes impliqués dans la voie de la réparation de l'ADN. Comme exemple, il a été démontré que les mutations monoalléliques de BRCA1, BRCA2, PALB2 et de MUTYH peuvent conduire à une instabilité génomique et à des défauts de réponse aux lésions de l'ADN, suggérant que l'haplo-insuffisance joue un rôle dans la stabilité génomique et la carcinogénèse (Arnold et al., 2006; Nikkilä et al., 2013; Grasso et al., 2014; Vaclová et al., 2015).

Notre stratégie dans ce travail était d'appliquer un stress génotoxique pour exacerber les différences existantes entre les cellules mutées et les cellules sauvages pour mettre en évidence un endophénotype typique des cellules mutées. Cette stratégie a été appliquée avec succès au gène TP53 par notre équipe (Zerdoumi et al., 2017). Les résultats obtenus lors de l'analyse supervisée des données globales de RNA-Seq montrent clairement une séparation nette des groupes de témoins et de mutants en présence d'un stress génotoxique. L'analyse globale du transcriptome des lymphocytes des patients représente donc un outil puissant pour prédire l'existence d'une mutation délétère du gène TP53. Compte-tenu de la puissance de résolution de cette méthode, il est fort probable qu'en augmentant le nombre de mutations testées, il soit possible de distinguer les différents types de mutations et peut-être leur impact différent sur les différentes fonctions du gène TP53. De façon intéressante, l'analyse des données globales du RNA-seq laisse penser qu'une différence existerait déjà à l'état basal sans traitement génotoxique entre les cellules sauvages et mutées pour TP53. Si ces données sont très préliminaires, l'accumulation de données de RNA-Seq à l'état basal d'un plus grand nombre de mutations devrait nous permettre de nous affranchir du bruit de fond de la variabilité d'expression inter-individus et nous permettre de vérifier cette observation.

Les données de RNA-Seq obtenues pour les mutants des gènes BRCA suggèrent également la possibilité de détecter un défaut d'un de ces deux gènes par une analyse transcriptomique. Ces résultats vont dans le sens de ceux obtenus par d'autres équipes qui ont

irradié des fibroblastes ou des cellules épithéliales mammaires du sein et de l'ovaire mutées à l'état hétérozygote pour les gènes BRCA1 et/ou BRCA2 et qui ont observé des différences d'expression liées au statut mutationnel (Kote-Jarai et al., 2004, 2006). Cependant, pour avoir une signature spécifique d'une altération d'un de ces deux gènes, le nombre de mutations testées doit absolument être augmenté. Il serait intéressant dans ce cas, d'essayer de discriminer également les groupes de mutants BRCA1 et BRCA2. La même méthode pourrait être appliquée pour identifier une signature spécifique d'autres gènes de prédisposition au cancer du sein et de l'ovaire comme PALB2 ou RAD51C/D. De façon intéressante, Pouliot et ses collaborateurs ont récemment publié un article tentant d'établir sur la base d'analyses RNA-Seq sur 117 lignées lymphoblastoïdes une signature différenciant les porteuses de mutations BRCA1/2 des femmes affectées ou non affectées à haut-risque de cancer non-lié à BRCA1 ou 2 (BRCA2) (Pouliot et al., 2017). Plus globalement, isoler une signature transcriptomique permettant d'identifier un défaut dans la voie de réparation des cassures double brin de l'ADN par recombinaison homologue serait une avancée majeure dans le contexte de l'AMM obtenue pour les inhibiteurs de PARP dans le traitement des cancers de l'ovaire et en prévision de celle qui pourrait être obtenue pour le cancer du sein.

L'utilisation de méthodes omiques pour appréhender les différences d'endophénotype entre les cellules sauvages et mutées à l'état hétérozygote semble donc être une approche prometteuse. Nous avons voulu dans ce projet évaluer la pertinence d'une approche à haut-débit du métabolome pour répondre à cette question. Récemment, une étude multi-omique visant à définir l'implication de p53 dans la régulation du métabolisme a été publiée (Huang et al., 2018). Cette étude basée sur l'intégration d'une analyse transcriptomique et d'une analyse métabolomique a été réalisée sur des cellules HCT116 de cancer colique humaines dépourvues de p53 et a mis en évidence le rôle joué par p53 dans le métabolisme des nucléotides et du soufre. Cependant, à notre connaissance, aucune étude multi-omique de ce genre n'a été rapportée pour fournir un profil global des cellules non tumorales portant une altération à l'état hétérozygote dans les gènes TP53 ou BRCA. Dans ce projet, l'analyse globale des métabolites polaires ne nous a pas permis dans les conditions utilisées de discriminer les groupes mutants des groupes sauvages et ceci même en présence d'un traitement génotoxique. Nous envisageons la possibilité dans les prochains mois d'analyser les mêmes échantillons avec des conditions de séparation différentes permettant de visualiser une fenêtre différente de métabolites. Pour chaque culot cellulaire isolé pour les analyses de métabolites, nous avons conservé le surnageant. Nous envisageons maintenant d'analyser le métabolome de ces surnageants. En effet, il semble exister à l'état basal des différences de

sécrétome entre les cellules sauvages et mutées à l'état hétérozygote pour TP53 (Pr David Malkin, données non publiées). Ces différences d'abondance de métabolites pourraient représenter de bons biomarqueurs de la présence d'une mutation délétère.

Le deuxième objectif de ce travail était d'extraire des données omiques les biomarqueurs les plus discriminants pour les intégrer dans des tests fonctionnels simples plus applicables en routine diagnostique. A partir des données de RNA-Seq obtenues pour les lignées TP53, nous avons développé un test ciblé basé sur l'utilisation de la RT-MPLA. Cette méthode simple et flexible permet d'effectuer des tests allant d'un à 40 échantillons en parallèle, ce qui la rend particulièrement adaptée à un flux de travail de diagnostic de routine. De plus, la procédure de RT-MLPA peut être réalisée en moins d'un jour avec un coût réactif de moins de 5 euros par échantillon. Elle permet de quantifier avec précision et rapidité l'expression relative de 20 transcrits d'intérêt dans un test à un tube en une seule réaction. De plus, cette méthode peut donner des informations quantitatives fiables même si la quantité de départ d'ARN est faible (Eldering et al., 2003). Ce nouvel essai, interrogeant plus de gènes cibles de p53 et permettant de monitorer de façon simultanée la quantité d'ARN de TP53, a été validé avec succès sur 34 lignées lymphoblastoïdes issues de patients Li-Fraumeni.

Nous avons dans un deuxième temps adapté cet essai pour qu'il puisse être réalisé directement sur un prélèvement sanguin. En effet, même si l'utilisation de lymphocytes immortalisés confère de nombreux avantages (ressources biologiques illimitées, facilité dans la maintenance des cellules en culture, temps de division rapide), le temps d'établissement d'une lignée cellulaire devient de moins en moins compatible avec les exigences diagnostiques. Ainsi, la possibilité de travailler sur un simple prélèvement sanguin offre une alternative peu invasive à l'établissement d'une lignée. Ce prélèvement pourrait par exemple être envisagé lors de la confirmation sur un deuxième prélèvement de la mutation détectée par NGS, permettant ainsi de ne pas multiplier les gestes médicaux sur les patients. Si les cellules sanguines ne représentent pas le meilleur modèle pour les études visant à mieux comprendre la physiopathologie de ces prédispositions génétiques au cancer, elles sont totalement appropriées pour notre étude qui vise à identifier des biomarqueurs sur des échantillons biologiques facilement accessibles. Les résultats obtenus sur deux témoins montrent la faisabilité de ce test. Nous allons maintenant valider son utilisation sur des prélèvements d'apparentés de porteurs de mutations délétères de TP53. Si les apparentés testés sont porteurs de la mutation, le test devrait présenter des faibles scores d'induction. Si ces apparentés sont

Objectif 2

non porteurs, le prélèvement devrait se comporter comme celui d'un individu témoin. En parallèle, nous allons devoir appréhender l'impact de facteurs possiblement confondants comme la présence de cellules leucémiques, de traitements de chimiothérapies ou encore l'âge des patients. Nous évaluerons dans les prochaines semaines l'impact d'une congélation des PBMC avant la mise en culture sur la réponse p53. Une réflexion est à mener quant aux témoins positifs les plus adaptés à ce test. Si la congélation était possible, il serait envisageable d'isoler les PBMC en masse de poches de sang de donneurs sains et de les congeler pour servir de témoins positifs d'induction de la voie p53. Nous envisageons également la possibilité d'intégrer un marqueur d'induction à nos réactions de RT-MLPA.

Conclusion

Aujourd'hui grâce aux avancées technologiques et au développement du séquençage à haut-débit, nous sommes passés à l'ère Génomique. Cette méthode révolutionnaire appliquée à la médecine a rendu possible le séquençage de milliers de génomes de patients. Le développement de techniques d'enrichissement et la miniaturisation des séquenceurs NGS ont permis une **entrée en masse de cette technologie dans les laboratoires de biologie médicale** des CHU et autres structures hospitalières privées participant au service public de santé en France. Ainsi, le séquençage ciblé par NGS de quelques dizaines à plusieurs centaines de gènes a rapidement supplanté le séquençage Sanger traditionnel et est devenu la norme expérimentale du diagnostic moléculaire pour les prédispositions génétiques au cancer. Cette évolution inéluctable conduit à la caractérisation **d'un nombre grandissant de variants** dont l'interprétation clinique est encore complexifiée par la découverte récente de la très grande diversité inter-individuelle du génome humain. Cette évolution des pratiques diagnostiques s'accompagne également d'une augmentation des **découvertes incidentelles de variants génétiques** dans des gènes hors spectre indiqué ou hors expertise du laboratoire. Les biologistes producteurs de données de séquençage se trouvent donc devant une difficulté supplémentaire qui nécessite de repenser les pratiques actuelles. L'enjeu de la génomique humaine auquel doivent faire face les médecins et généticiens n'est plus, compte-tenu de cette évolution technologique, la détection des altérations mais leur interprétation. Ainsi, pour que ces avancées technologiques puissent se traduire en bénéfice pour les patients, il est indispensable de **développer des stratégies d'interprétation des variants** dans lesquels les tests fonctionnels prendront vraisemblablement une place grandissante.

L'essai fonctionnel dédié à TP53 que nous avons développé permettra **d'apporter des arguments forts pour classer un variant** en variant bénin ou pathogène. La possibilité d'effectuer ce test sur prélèvement sanguin en moins d'une semaine rend ce test **compatible avec les délais diagnostiques**. En effet, aujourd'hui, **l'identification chez un patient d'une mutation délétère du gène TP53 va bouleverser sa prise en charge médicale** et un protocole de surveillance lourd sera mis en place dans le but de dépister précocement les tumeurs (Protocole de Toronto, (Villani et al., 2016)). De plus, pour les patients Li-Fraumeni, il est recommandé d'éviter l'exposition aux rayons. En effet, compte-tenu du nombre important de cancers secondaires développés dans la zone d'irradiation, il est recommandé dans la mesure du possible d'éviter la radiothérapie chez ces patients (Salmon et al., 2007; Heymann et al., 2010). Aujourd'hui, il est recommandé de surveiller le développement tumoral chez les patients Li-Fraumeni par IRM corps entier, mammaire et cérébral annuel. Cette technique ne faisant intervenir aucun rayonnement ionisant mais un champ

électromagnétique, elle semble être dépourvue d'effet génotoxique. La détection chez un patient d'une mutation délétère de TP53 pourrait également dans le futur orienter le choix de la chimiothérapie à utiliser. Une étude récente menée au sein de notre équipe a confirmé que la plupart des chimiothérapies administrées au cours du premier cancer déclenchent des cancers secondaires par leur effet génotoxique à l'exception des poisons du fuseau (Kasper et al., 2018) (**Article Annexe 1**).

La plupart des tests fonctionnels disponibles pour des gènes aux fonctions multiples comme les gènes BRCA font intervenir des étapes de clonage, de mutagenèse dirigée de transfection et de culture cellulaire et sont souvent spécifiques de domaines fonctionnels. Ces tests très intéressants pour **l'interprétation des VSI** sont cependant peu compatibles avec la chronologie du diagnostic moléculaire à l'ère de la génomique et de la médecine personnalisée où la découverte d'un variant chez un patient peut conditionner sa prise en charge immédiate. Si des tests, comme celui que nous avons développé, réalisables directement sur prélèvement sanguin en moins d'une semaine, représentent une réponse potentielle à cette problématique, le **développement de stratégies à haut-débit** pour tester la fonctionnalité de toutes les altérations possibles pouvant toucher un gène est une alternative très intéressante (Starita et al., 2018; Findlay et al., 2018). A terme, on peut penser que toutes les substitutions possibles des gènes impliqués dans les prédispositions génétiques au cancer pourront être recensées dans des bases de données avec un score de fonctionnalité qui pourra être intégré automatiquement à l'annotation des variants détectés en NGS. Ces approches à haut-débit représentent une évolution majeure dans le domaine des tests fonctionnels. Cependant, ces tests sont réalisés indépendamment du contexte génétique du patient et des approches comme celles proposées dans ce travail de thèse pourront venir compléter l'information apportée par ces tests.

Il est également possible d'utiliser les tests fonctionnels que nous avons développés dans le but de **détecter la présence de mutations cryptiques** impactant la fonction du gène étudié. Ces altérations peuvent correspondre à des mutations introniques profondes, à des mutations des régions régulatrices proximales ou distales, ou encore à des altérations épigénétiques qui peuvent échapper au criblage mutationnel réalisé au titre du diagnostic. Devant une présentation clinique très évocatrice, un test fonctionnel pourra être demandé même si l'analyse mutationnelle n'a pas révélé de variant délétère. Ceci constituera un argument en faveur de l'implication du gène dans le phénotype et pourra motiver la recherche des altérations cryptiques en cause.

Comme notre équipe l'a démontré dans la première partie de ce travail, les tests fonctionnels dédiés à un gène trouvent également leur place pour affiner les connaissances concernant la **relation entre le génotype et le phénotype**. En effet, il existe une grande hétérogénéité des présentations phénotypiques. Cette hétérogénéité peut concerner l'âge de survenue des tumeurs, la localisation tumorale ou encore le nombre de tumeurs primitives que pourra développer un individu. La nature et la localisation des mutations impacte vraisemblablement cette hétérogénéité. Cette hétérogénéité phénotypique représente un obstacle majeur à la prise en charge médicale adaptée des patients présentant une mutation constitutionnelle de *TP53*. En effet, dans le LFS et plus largement dans les pathologies cancéreuses liées aux mutations constitutionnelles de *TP53* (*p53-related cancers*), **l'évaluation du risque tumoral en fonction du type de mutation** est particulièrement difficile compte-tenu de la faible prévalence de ce syndrome et du biais de recrutement des patients, basé essentiellement sur l'agrégation familiale et l'apparition précoce des tumeurs. Plusieurs équipes ont analysé le spectre tumoral associé aux mutations constitutionnelles de *TP53* afin d'établir des corrélations entre le génotype et le phénotype, tant au niveau du type de tumeur que de l'âge de survenue (Birch *et al.*, 1998; Monti *et al.*, 2007; Bougeard *et al.*, 2008, 2015). Ces études ont montré que les mutations faux-sens à effet dominant négatif sont associées à un âge de développement tumoral plus précoce et à une incidence plus élevée de cancer que les autres types de mutations. Grâce au test fonctionnel développé par notre équipe ainsi qu'aux analyses de ChIP-Seq réalisées, nous avons pu déterminer que ce gradient de sévérité clinique résultait d'un gradient de défaut de la réponse de p53 aux lésions de l'ADN lié à une altération de la fixation de la protéine p53 à l'ADN (Zerdoumi *et al.*, 2017). Nos travaux ont différentes implications potentielles sur la prise en charge médicale et thérapeutique des patients. L'évaluation de l'effet biologique réel des mutations hétérozygotes dans le contexte génétique du patient par ces tests fonctionnels va permettre **l'évaluation du risque de tumeur en fonction du type de mutation** et par la suite **d'adapter la prise en charge du patient en fonction de ce risque**. Le protocole de Toronto qui a pour but le dépistage précoce des tumeurs est très lourd pour le patient et sa famille (Villani *et al.*, 2016). Afin de limiter ces contraintes, le groupe français de travail sur le LFS préconise d'adapter le suivi au type de mutation identifiée. Dans ce contexte, les résultats de notre essai fonctionnel incitent les médecins à proposer une **stratification de la prise en charge du syndrome de Li-Fraumeni, en fonction du type de mutation identifiée chez le patient**. Dans les familles présentant une mutation associée à une forme sévère de la maladie, il est sans doute légitime de proposer l'analyse du gène *TP53* aux enfants, dès lors que les parents le souhaitent après

avoir reçu une information au cours d'une consultation de génétique. Les enfants porteurs de la mutation pourront bénéficier d'un **suivi annuel fondé sur l'IRM corps entier**. Un suivi dont il faudra du reste évaluer l'efficacité. Un diagnostic prénatal ou préimplantatoire pourra par ailleurs être proposé pour la descendance. Dans le cas de la détection d'une mutation associée aux formes moins sévères, l'analyse pré-symptomatique pourra être réservée aux adultes, **en limitant le dépistage des tumeurs à celui des cancers du sein par IRM annuelle, à partir de 20 ans**. L'objectif est d'améliorer la prise en charge des familles, en intégrant l'hétérogénéité des mutations et la gravité de ce syndrome, et en considérant le **rapport bénéfice médical/impact psychologique**.

Les prédispositions génétiques au cancer liées à des altérations constitutionnelles de gènes suppresseurs de tumeurs sont généralement associées à **une pénétrance incomplète**. Dans ce contexte, identifier les individus porteurs de la mutation n'est plus la seule question et le challenge consistera à déterminer quel apparenté porteur de la mutation développera un cancer. Dans ce contexte, les approches multi-omiques réalisées sur des apparentés porteurs indemnes et des apparentés porteurs ayant développé une tumeur semblent constituer des approches intéressantes pour identifier des biomarqueurs pouvant être utilisés pour **prédire un risque de développement tumoral**. Dans cette même optique, des tests fonctionnels réalisés dans le contexte génétique des patients en combinaison avec des analyses d'exomes ou de génomes peuvent constituer des outils puissants pour identifier des **facteurs génétiques modificateurs** de la sévérité du phénotype. Par exemple, si un score de fonctionnalité très différent est obtenu pour deux patients porteurs de la même mutation et présentant des phénotypes de sévérité différente, un facteur modificateur pourra être recherché par séquençage exomique. Notre équipe s'est récemment investie dans la recherche de facteurs génétiques modificateurs de l'impact de mutations délétères de *TP53* par une approche de séquençage exomique de duos composés d'enfants ayant développé une tumeur pédiatrique et de leur parent porteur de la mutation indemne de cancer ou ayant présenté une tumeur à un âge plus tardif. Les tests fonctionnels *TP53* développés pourront venir compléter ces investigations.

Les tests fonctionnels spécifiques de gènes ou de voies biologiques ont donc toute leur place à l'ère de la génomique et de la médecine personnalisée et représentent donc un enjeu majeur qui permettra de décoder toutes les données de séquençage issues de patients ou d'individus témoins. Les **méthodes omiques** telles que celles utilisées dans ce travail

représentent **des outils puissants pour identifier des signatures associées à un phénotype**. Elles représentent également **un moyen d'augmenter le débit des tests fonctionnels** qui jusqu'alors ne présentent pas des délais compatibles avec le diagnostic. Plus largement, avec l'évolution et la démocratisation des technologies à haut-débit, il est possible que dans un futur proche, pour chaque patient, les analyses génomiques soient complétées par des analyses épigénétiques, transcriptomiques, protéomiques ou métabolomiques permettant de s'approcher au plus près du phénotype du patient. Ces futurs développements laissent envisager la possibilité de développer pour chaque patient, **une médecine personnalisée, de précision dans laquelle l'information génétique pourra être interprétée en lumière des signatures biologiques identifiées**.

Références bibliographiques

1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.

Abramova NA, Russell J, Botchan M, Li R. 1997. Interaction between replication protein A and p53 is disrupted after UV damage in a DNA repair-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 7186–7191.

Adamo A, Collis SJ, Adelman CA, Silva N, Horejsi Z, Ward JD, Martinez-Perez E, Boulton SJ, La Volpe A. 2010. Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Molecular Cell* **39**: 25–35.

Adimoolam S, Ford JM. 2002. p53 and DNA damage-inducible expression of the xeroderma pigmentosum group C gene. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 12985–12990.

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics* **Chapter 7**: Unit7.20.

Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nature Reviews. Genetics* **12**: 363–376.

Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, Berg JS, Biswas S, Bowling KM, Conlin LK, et al. 2016. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *American Journal of Human Genetics* **99**: 247.

Andrade KC, Mirabello L, Stewart DR, Karlins E, Koster R, Wang M, Gapstur SM, Gaudet MM, Freedman ND, Landi MT, et al. 2017. Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. *Human Mutation* **38**: 1723–1730.

Andreasen PA, Egelund R, Petersen HH. 2000. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cellular and molecular life sciences: CMLS* **57**: 25–40.

Aramayo R, Sherman MB, Brownless K, Lurz R, Okorokov AL, Orlova EV. 2011. Quaternary structure of the specific p53-DNA complex reveals the mechanism of p53 mutant dominance. *Nucleic Acids Research* **39**: 8960–8971.

Arias-Lopez C, Lazaro-Trueba I, Kerr P, Lord CJ, Dexter T, Irvani M, Ashworth A, Silva A. 2006. p53 modulates homologous recombination by transcriptional regulation of the RAD51 gene. *EMBO reports* **7**: 219–224.

Arnold K, Kim M-K, Frerk K, Edler L, Savelyeva L, Schmezer P, Wiedemeyer R. 2006. Lower level of BRCA2 protein in heterozygous mutation carriers is correlated with an increase in DNA double strand breaks and an impaired DSB repair. *Cancer Letters* **243**: 90–100.

Aubrey BJ, Kelly GL, Janic A, Herold MJ, Strasser A. 2018. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation* **25**: 104–113.

- Baert-Desurmont S, Coutant S, Charbonnier F, Macquere P, Lecoquierre F, Schwartz M, Blanluet M, Vezain M, Lanos R, Quenez O, et al. 2018.** Optimization of the diagnosis of inherited colorectal cancer using NGS and capture of exonic and intronic sequences of panel genes. *European journal of human genetics: EJHG* **26**: 1597–1602.
- Bahassi EM, Stambrook PJ. 2014.** Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* **29**: 303–310.
- Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, vanTuinen P, Ledbetter DH, Barker DF, Nakamura Y, et al. 1989.** Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science (New York, N.Y.)* **244**: 217–221.
- Balog J, Sasi-Szabó L, Kinross J, Lewis MR, Muirhead LJ, Veselkov K, Mirnezami R, Dezsó B, Damjanovich L, Darzi A, et al. 2013.** Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Science Translational Medicine* **5**: 194ra93.
- Baralle D, Lucassen A, Buratti E. 2009.** Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO reports* **10**: 810–816.
- Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR. 2017.** Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods* **14**: 135–139.
- Bashiardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M. 2005.** Direct genomic selection. *Nature Methods* **2**: 63–69.
- Beckerman R, Prives C. 2010.** Transcriptional regulation by p53. *Cold Spring Harbor Perspectives in Biology* **2**: a000935.
- Benton HP, Want E, Keun HC, Amberg A, Plumb RS, Goldfain-Blanc F, Walther B, Reily MD, Lindon JC, Holmes E, et al. 2012.** Intra- and interlaboratory reproducibility of ultra performance liquid chromatography-time-of-flight mass spectrometry for urinary metabolic profiling. *Analytical Chemistry* **84**: 2424–2432.
- Bergers G, Benjamin LE. 2003.** Tumorigenesis and the angiogenic switch. *Nature Reviews. Cancer* **3**: 401–410.
- Béroud C, Letovsky SI, Braastad CD, Caputo SM, Beaudoux O, Bignon YJ, Bressac-De Paillerets B, Bronner M, Buell CM, Collod-Béroud G, et al. 2016.** BRCA Share: A Collection of Clinical BRCA Gene Variants. *Human Mutation* **37**: 1318–1328.
- Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014.** BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**: 362–365.
- Bialik S, Dasari SK, Kimchi A. 2018.** Autophagy-dependent cell death - where, how and why a cell eats itself to death. *Journal of Cell Science* **131**.
- Birch JM, Blair V, Kelsey AM, Evans DG, Harris M, Tricker KJ, Varley JM. 1998.** Cancer phenotype correlates with constitutional TP53 genotype in families with the Li-Fraumeni syndrome. *Oncogene* **17**: 1061–1068.

- Bobée V, Ruminy P, Marchand V, Viailly P-J, Abdel Sater A, Veresezan L, Drioux F, Bérard C, Bohers E, Mareschal S, et al. 2017.** Determination of Molecular Subtypes of Diffuse Large B-Cell Lymphoma Using a Reverse Transcriptase Multiplex Ligation-Dependent Probe Amplification Classifier: A CALYM Study. *The Journal of molecular diagnostics: JMD* **19**: 892–904.
- Boland JF, Chung CC, Roberson D, Mitchell J, Zhang X, Im KM, He J, Chanock SJ, Yeager M, Dean M. 2013.** The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Human Genetics* **132**: 1153–1163.
- Bonadona V, Bonaïti B, Olschwang S, Grandjouan S, Huiart L, Longy M, Guimbaud R, Buecher B, Bignon Y-J, Caron O, et al. 2011.** Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* **305**: 2304–2310.
- Bork P, Blomberg N, Nilges M. 1996.** Internal repeats in the BRCA2 protein sequence. *Nature Genetics* **13**: 22–23.
- Bougeard G, Brugières L, Chompret A, Gesta P, Charbonnier F, Valent A, Martin C, Raux G, Feunteun J, Bressac-de Paillerets B, et al. 2003.** Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. *Oncogene* **22**: 840–846.
- Bougeard G, Renaux-Petel M, Flaman J-M, Charbonnier C, Fermey P, Belotti M, Gauthier-Villars M, Stoppa-Lyonnet D, Consolino E, Brugières L, et al. 2015.** Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **33**: 2345–2352.
- Bougeard G, Sesboüé R, Baert-Desurmont S, Vasseur S, Martin C, Tinat J, Brugières L, Chompret A, de Paillerets BB, Stoppa-Lyonnet D, et al. 2008.** Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *Journal of Medical Genetics* **45**: 535–538.
- Bouvard V, Zaitchouk T, Vacher M, Duthu A, Canivet M, Choisy-Rossi C, Nieruchalski M, May E. 2000.** Tissue and cell-specific expression of the p53-target genes: bax, fas, mdm2 and waf1/p21, before and following ionising irradiation in mice. *Oncogene* **19**: 649–660.
- Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P, Pieterse M, Wientjens E, Seibler J, Hogervorst FBL, et al. 2013.** A high-throughput functional complementation assay for classification of BRCA1 missense variants. *Cancer Discovery* **3**: 1142–1155.
- Boyle JM, Mitchell EL, Greaves MJ, Roberts SA, Tricker K, Burt E, Varley JM, Birch JM, Scott D. 1998.** Chromosome instability is a predominant trait of fibroblasts from Li-Fraumeni families. *British Journal of Cancer* **77**: 2181–2192.
- Brosh R, Rotter V. 2009.** When mutants gain new powers: news from the mutant p53 field. *Nature Reviews. Cancer* **9**: 701–713.
- Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ, Helleday T. 2005.** Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**: 913–917.

- Bubien V, Bonnet F, Brouste V, Hoppe S, Barouk-Simonet E, David A, Edery P, Bottani A, Layet V, Caron O, et al. 2013.** High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. *Journal of Medical Genetics* **50**: 255–263.
- Budanov AV. 2011.** Stress-Responsive Sestrins Link p53 with Redox Regulation and Mammalian Target of Rapamycin Signaling. *Antioxidants & Redox Signaling* **15**: 1679–1690.
- Budanov AV, Karin M. 2008.** p53 target genes sestrin1 and sestrin2 connect genotoxic stress and mTOR signaling. *Cell* **134**: 451–460.
- Bullock AN, Fersht AR. 2001.** Rescuing the function of mutant p53. *Nature Reviews. Cancer* **1**: 68–76.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011.** Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development* **25**: 1915–1927.
- Camplejohn RS, Perry P, Hodgson SV, Turner G, Williams A, Upton C, MacGeoch C, Mohammed S, Barnes DM. 1995.** A possible screening test for inherited p53-related defects based on the apoptotic response of peripheral blood lymphocytes to DNA damage. *British Journal of Cancer* **72**: 654–662.
- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, et al. 2001.** BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**: 149–160.
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003.** ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Research* **31**: 3568–3571.
- Castéra L, Harter V, Muller E, Krieger S, Goardon N, Ricou A, Rousselin A, Paimparay G, Legros A, Bruet O, et al. 2018.** Landscape of pathogenic variations in a panel of 34 genes and cancer risk estimation from 5131 HBOC families. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*.
- Castéra L, Krieger S, Rousselin A, Legros A, Baumann J-J, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, et al. 2014.** Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *European journal of human genetics: EJHG* **22**: 1305–1313.
- Chang S, Biswas K, Martin BK, Stauffer S, Sharan SK. 2009.** Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. *The Journal of Clinical Investigation* **119**: 3160–3171.
- Chang C-J, Chao C-H, Xia W, Yang J-Y, Xiong Y, Li C-W, Yu W-H, Rehman SK, Hsu JL, Lee H-H, et al. 2011.** p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nature Cell Biology* **13**: 317–323.
- Chatterjee N, Walker GC. 2017.** Mechanisms of DNA damage, repair and mutagenesis. *Environmental and molecular mutagenesis* **58**: 235–263.

- Chen C-C, Feng W, Lim PX, Kass EM, Jasin M. 2018.** Homology-Directed Repair and the Role of BRCA1, BRCA2, and Related Proteins in Genome Integrity and Cancer. *Annual Review of Cancer Biology* **2**: 313–336.
- Chen IT, Smith ML, O'Connor PM, Fornace AJ. 1995.** Direct interaction of Gadd45 with PCNA and evidence for competitive interaction of Gadd45 and p21Waf1/Cip1 with PCNA. *Oncogene* **11**: 1931–1937.
- Ching T, Huang S, Garmire LX. 2014.** Power analysis and sample size estimation for RNA-Seq differential expression. *RNA (New York, N.Y.)* **20**: 1684–1696.
- Chipuk JE, Kuwana T, Bouchier-Hayes L, Droin NM, Newmeyer DD, Schuler M, Green DR. 2004.** Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science (New York, N.Y.)* **303**: 1010–1014.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, et al. 2009.** Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 19096–19101.
- Chompret A, Abel A, Stoppa-Lyonnet D, Brugières L, Pagés S, Feunteun J, Bonaïti-Pellié C. 2001.** Sensitivity and predictive value of criteria for p53 germline mutation screening. *Journal of Medical Genetics* **38**: 43–47.
- Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H. 2009.** Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* **4**: 265–270.
- Claus EB, Risch N, Thompson WD. 1991.** Genetic analysis of breast cancer in the cancer and steroid hormone study. *American Journal of Human Genetics* **48**: 232–242.
- Cochran RL, Cidado J, Kim M, Zabransky DJ, Croessmann S, Chu D, Wong HY, Beaver JA, Cravero K, Erlanger B, et al. 2015.** Functional isogenic modeling of BRCA1 alleles reveals distinct carrier phenotypes. *Oncotarget* **6**: 25240–25251.
- Crichton D, Wilkinson S, O'Prey J, Syed N, Smith P, Harrison PR, Gasco M, Garrone O, Crook T, Ryan KM. 2006.** DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell* **126**: 121–134.
- Crouse GF. 2016.** Non-canonical actions of mismatch repair. *DNA repair* **38**: 102–109.
- Daniels MJ, Wang Y, Lee M, Venkitaraman AR. 2004.** Abnormal cytokinesis in cells deficient in the breast cancer susceptibility protein BRCA2. *Science (New York, N.Y.)* **306**: 876–879.
- De Braekeleer M, De Braekeleer E, Douet-Guilbert N. 2015.** Whole-Genome/Exome Sequencing in Acute Leukemia: From Research to Clinics. In: Wu W, Choudhry H, eds. *Next Generation Sequencing in Cancer Research, Volume 2: From Basepairs to Bedsides*. Cham: Springer International Publishing, 381–400.
- Deans AJ, West SC. 2011.** DNA interstrand crosslink repair and cancer. *Nature Reviews. Cancer* **11**: 467–480.

- el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. 1992.** Definition of a consensus binding site for p53. *Nature Genetics* **1**: 45–49.
- Densham RM, Garvin AJ, Stone HR, Strachan J, Baldock RA, Daza-Martin M, Fletcher A, Blair-Reid S, Beesley J, Johal B, et al. 2016.** Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nature Structural & Molecular Biology* **23**: 647–655.
- Desmet F-O, Hamroun D, Lalande M, Collod-Bérourd G, Claustres M, Bérourd C. 2009.** Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* **37**: e67.
- Dianov GL, Hübscher U. 2013.** Mammalian Base Excision Repair: the Forgotten Archangel. *Nucleic Acids Research* **41**: 3483–3490.
- Dittmer D, Pati S, Zambetti G, Chu S, Teresky AK, Moore M, Finlay C, Levine AJ. 1993.** Gain of function mutations in p53. *Nature Genetics* **4**: 42–46.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**: 15–21.
- Doxsey SJ, Stein P, Evans L, Calarco PD, Kirschner M. 1994.** Pericentrin, a highly conserved centrosome protein involved in microtubule organization. *Cell* **76**: 639–650.
- Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. 2011.** Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews* **40**: 387–426.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009.** Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)* **323**: 133–138.
- Eldering E, Spek CA, Aberson HL, Grummels A, Derks IA, de Vos AF, McElgunn CJ, Schouten JP. 2003.** Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Research* **31**: e153.
- Eliyahu D, Raz A, Gruss P, Givol D, Oren M. 1984.** Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. *Nature* **312**: 646–649.
- ENCODE Project Consortium. 2012.** An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
- Farmer G, Bargonetti J, Zhu H, Friedman P, Prywes R, Prives C. 1992.** Wild-type p53 activates transcription in vitro. *Nature* **358**: 83–86.
- Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, et al. 2005.** Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**: 917–921.

- Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian SV, et al. 2008.** Functional Assays for Classification of BRCA2 Variants of Uncertain Significance. *Cancer research* **68**: 3523–3531.
- Fiehn O. 2002.** Metabolomics--the link between genotypes and phenotypes. *Plant Molecular Biology* **48**: 155–171.
- Fields S, Jang SK. 1990.** Presence of a potent transcription activating sequence in the p53 protein. *Science (New York, N.Y.)* **249**: 1046–1049.
- Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. 2018.** Accurate classification of BRCA1 variants with saturation genome editing. *Nature*.
- Finlay CA, Hinds PW, Levine AJ. 1989.** The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**: 1083–1093.
- Fischer M. 2017.** Census and evaluation of p53 target genes. *Oncogene* **36**: 3943–3956.
- Flaman JM, Frebourg T, Moreau V, Charbonnier F, Martin C, Chappuis P, Sappino AP, Limacher IM, Bron L, Benhattar J. 1995.** A simple p53 functional assay for screening cell lines, blood, and tumors. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 3963–3967.
- Flaman JM, Robert V, Lenglet S, Moreau V, Iggo R, Frebourg T. 1998.** Identification of human p53 mutations with differential effects on the bax and p21 promoters using functional assays in yeast. *Oncogene* **16**: 1369–1372.
- Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. 2011.** LOVD v.2.0: the next generation in gene variant databases. *Human Mutation* **32**: 557–563.
- Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, et al. 2017.** COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research* **45**: D777–D783.
- Fouad YA, Aanei C. 2017.** Revisiting the hallmarks of cancer. *American Journal of Cancer Research* **7**: 1016–1036.
- Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, Genome of the Netherlands Consortium, van Duijn CM, Swertz M, Wijmenga C, et al. 2015.** Genome-wide patterns and properties of de novo mutations in humans. *Nature Genetics* **47**: 822–826.
- Frebourg T, Barbier N, Kassel J, Ng YS, Romero P, Friend SH. 1992.** A functional screen for germ line p53 mutations based on transcriptional activation. *Cancer Research* **52**: 6976–6978.
- Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. 2010.** Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods in Molecular Biology (Clifton, N.J.)* **653**: 249–257.

- Gee ME, Faraahi Z, McCormick A, Edmondson RJ. 2018.** DNA damage repair in ovarian cancer: unlocking the heterogeneity. *Journal of Ovarian Research* **11**: 50.
- Goi K, Takagi M, Iwata S, Delia D, Asada M, Donghi R, Tsunematsu Y, Nakazawa S, Yamamoto H, Yokota J, et al. 1997.** DNA damage-associated dysregulation of the cell cycle and apoptosis control in cells with germ-line p53 mutation. *Cancer Research* **57**: 1895–1902.
- Gong Z, Kim J-E, Leung CCY, Glover JNM, Chen J. 2010.** BACH1/FANCI acts with TopBP1 and participates early in DNA replication checkpoint control. *Molecular Cell* **37**: 438–446.
- González-Pérez A, López-Bigas N. 2011.** Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics* **88**: 440–449.
- Goodwin S, McPherson JD, McCombie WR. 2016.** Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews. Genetics* **17**: 333–351.
- Gowen LC, Johnson BL, Latour AM, Sulik KK, Koller BH. 1996.** Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nature Genetics* **12**: 191–194.
- Grasso F, Giacomini E, Sanchez M, Degan P, Gismondi V, Mazzei F, Varesco L, Viel A, Bignami M. 2014.** Genetic instability in lymphoblastoid cell lines expressing biallelic and monoallelic variants in the human MUTYH gene. *Human Molecular Genetics* **23**: 3843–3852.
- Guidugli L, Carreira A, Caputo SM, Ehlen A, Galli A, Monteiro ANA, Neuhausen SL, Hansen TVO, Couch FJ, Vreeswijk MPG, et al. 2014.** Functional assays for analysis of variants of uncertain significance in BRCA2. *Human Mutation* **35**: 151–164.
- Guidugli L, Pankratz VS, Singh N, Thompson J, Erding CA, Engel C, Schmutzler R, Domchek S, Nathanson K, Radice P, et al. 2013.** A classification model for BRCA2 DNA binding domain missense variants based on homology directed repair activity. *Cancer research* **73**: 265–275.
- Hainaut P, Hollstein M. 2000.** p53 and human cancer: the first ten thousand mutations. *Advances in Cancer Research* **77**: 81–137.
- Hakem R, de la Pompa JL, Sirard C, Mo R, Woo M, Hakem A, Wakeham A, Potter J, Reitmair A, Billia F, et al. 1996.** The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. *Cell* **85**: 1009–1023.
- Han X, Saito H, Miki Y, Nakanishi A. 2008.** A CRM1-mediated nuclear export signal governs cytoplasmic localization of BRCA2 and is essential for centrosomal localization of BRCA2. *Oncogene* **27**: 2969–2977.
- Hanahan D, Weinberg RA. 2000.** The hallmarks of cancer. *Cell* **100**: 57–70.
- Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, Schrader KA, Schaeffer DF, Shumansky K, Zogopoulos G, et al. 2015.** Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA oncology* **1**: 23–32.

- Harkin DP, Bean JM, Miklos D, Song YH, Truong VB, Englert C, Christians FC, Ellisen LW, Maheswaran S, Oliner JD, et al. 1999.** Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97**: 575–586.
- Hartmann L, Theiss S, Niederacher D, Schaal H. 2008.** Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Frontiers in Bioscience: A Journal and Virtual Library* **13**: 3252–3272.
- Hasty P, Sharp ZD, Curiel TJ, Campisi J. 2013.** mTORC1 and p53. *Cell Cycle* **12**: 20–25.
- Haupt Y, Maya R, Kazaz A, Oren M. 1997.** Mdm2 promotes the rapid degradation of p53. *Nature* **387**: 296–299.
- Hermeking H, Lengauer C, Polyak K, He TC, Zhang L, Thiagalingam S, Kinzler KW, Vogelstein B. 1997.** 14-3-3sigma is a p53-regulated inhibitor of G2/M progression. *Molecular Cell* **1**: 3–11.
- Heymann S, Delalogue S, Rahal A, Caron O, Frebourg T, Barreau L, Pachtet C, Mathieu M-C, Marsiglia H, Bourgier C. 2010.** Radio-induced malignancies after breast cancer postoperative radiotherapy in patients with Li-Fraumeni syndrome. *Radiation Oncology (London, England)* **5**: 104.
- Hinds PW, Finlay CA, Quartin RS, Baker SJ, Fearon ER, Vogelstein B, Levine AJ. 1990.** Mutant p53 DNA clones from human colon carcinomas cooperate with ras in transforming primary rat cells: a comparison of the ‘hot spot’ mutant phenotypes. *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research* **1**: 571–580.
- Hollis RL, Churchman M, Gourley C. 2017.** Distinct implications of different BRCA mutations: efficacy of cytotoxic chemotherapy, PARP inhibition and clinical outcome in ovarian cancer. *OncoTargets and therapy* **10**: 2539–2551.
- Holmes E, Wilson ID, Nicholson JK. 2008.** Metabolic phenotyping in health and disease. *Cell* **134**: 714–717.
- Honda R, Tanaka H, Yasuda H. 1997.** Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS letters* **420**: 25–27.
- Hong WS, Shpak M, Townsend JP. 2015.** Inferring the origin of metastases from cancer phylogenies. *Cancer research* **75**: 4021–4025.
- Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, et al. 2012.** Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation* **33**: 1228–1238.
- Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, et al. 2002.** Biallelic inactivation of BRCA2 in Fanconi anemia. *Science (New York, N.Y.)* **297**: 606–609.
- Huang J, Long Z, Lin W, Liao X, Xie Y, Liu L, Ma W. 2018.** Integrative omics analysis of p53-dependent regulation of metabolism. *FEBS letters* **592**: 380–393.

- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, et al. 2007.** DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Research* **35**: W169-175.
- Humpton TJ, Vousden KH. 2016.** Regulation of Cellular Metabolism and Hypoxia by p53. *Cold Spring Harbor Perspectives in Medicine* **6**.
- Imbriano C, Gurtner A, Cocchiarella F, Di Agostino S, Basile V, Gostissa M, Dobbstein M, Del Sal G, Piaggio G, Mantovani R. 2005.** Direct p53 transcriptional repression: in vivo analysis of CCAAT-containing G2/M promoters. *Molecular and Cellular Biology* **25**: 3737–3751.
- International Human Genome Sequencing Consortium. 2004.** Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931–945.
- Ishioka C, Frebourg T, Yan YX, Vidal M, Friend SH, Schmidt S, Iggo R. 1993.** Screening patients for heterozygous p53 mutations using a functional assay in yeast. *Nature Genetics* **5**: 124–129.
- Itahana Y, Itahana K. 2018.** Emerging Roles of p53 Family Members in Glucose Metabolism. *International Journal of Molecular Sciences* **19**.
- Jiang P, Du W, Wang X, Mancuso A, Gao X, Wu M, Yang X. 2011.** p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nature cell biology* **13**: 310–316.
- Jiang L, Kon N, Li T, Wang S-J, Su T, Hibshoosh H, Baer R, Gu W. 2015.** Ferroptosis as a p53-mediated activity during tumour suppression. *Nature* **520**: 57–62.
- Jiricny J. 2013.** Postreplicative Mismatch Repair. *Cold Spring Harbor Perspectives in Biology* **5**.
- Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, Lin JC-H, Palmisano E, Brune K, Jaffee EM, et al. 2009.** Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science (New York, N.Y.)* **324**: 217.
- Kanakis GA, Jørgensen N, Goulis DG. 2018.** Breast Cancer in Men. *The New England Journal of Medicine* **379**: 1385.
- Kasper E, Angot E, Colasse E, Nicol L, Sabourin J-C, Adriouch S, Lacoume Y, Charbonnier C, Raad S, Frebourg T, et al. 2018.** Contribution of genotoxic anticancer treatments to the development of multiple primary tumours in the context of germline TP53 mutations. *European Journal of Cancer (Oxford, England: 1990)* **101**: 254–262.
- Kastenhuber E, Lowe S. 2017.** Putting p53 in context. *Cell* **170**: 1062–1078.
- Katki HA, Greene MH, Achatz MI. 2018.** Testing Positive on a Multigene Panel Does Not Suffice to Determine Disease Risks. *Journal of the National Cancer Institute* **110**: 797–798.
- Kato S, Han S-Y, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C. 2003.** Understanding the function-structure and function-mutation relationships of p53 tumor

suppressor protein by high-resolution missense mutation analysis. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 8424–8429.

Kawauchi K, Araki K, Tobiume K, Tanaka N. 2008. p53 regulates glucose metabolism through an IKK-NF-kappaB pathway and inhibits cell transformation. *Nature Cell Biology* **10**: 611–618.

Khanduja JS, Calvo IA, Joh RI, Hill IT, Motamedi M. 2016. Nuclear Noncoding RNAs and Genome Stability. *Molecular Cell* **63**: 7–20.

Kim H, D'Andrea AD. 2012. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes & Development* **26**: 1393–1408.

Kim G, Ison G, McKee AE, Zhang H, Tang S, Gwise T, Sridhara R, Lee E, Tzou A, Philip R, et al. 2015a. FDA Approval Summary: Olaparib Monotherapy in Patients with Deleterious Germline BRCA-Mutated Advanced Ovarian Cancer Treated with Three or More Lines of Chemotherapy. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **21**: 4257–4261.

Kim NH, Kim HS, Li X-Y, Lee I, Choi H-S, Kang SE, Cha SY, Ryu JK, Yoon D, Fearon ER, et al. 2011. A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *The Journal of Cell Biology* **195**: 417–433.

Kim JW, Lyu SW, Sung SR, Sung SR, Park JE, Cha DH, Yoon TK, Ko JJ, Shim SH. 2015b. Molecular analysis of miscarriage products using multiplex ligation-dependent probe amplification (MLPA): alternative to conventional karyotype analysis. *Archives of Gynecology and Obstetrics* **291**: 347–354.

Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**: 310–315.

Knudson AG. 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* **68**: 820–823.

Kobayashi H, Ohno S, Sasaki Y, Matsuura M. 2013. Hereditary breast and ovarian cancer susceptibility genes (review). *Oncology Reports* **30**: 1019–1029.

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* **22**: 568–576.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature* **488**: 471–475.

Kote-Jarai Z, Matthews L, Osorio A, Shanley S, Giddings I, Moreews F, Locke I, Evans DG, Eccles D, Carrier Clinic Collaborators, et al. 2006. Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **12**: 3896–3901.

- Kote-Jarai Z, Williams RD, Cattini N, Copeland M, Giddings I, Wooster R, tePoele RH, Workman P, Gusterson B, Peacock J, et al. 2004.** Gene expression profiling after radiation-induced DNA damage is strongly predictive of BRCA1 mutation carrier status. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **10**: 958–963.
- Krizhanovsky V, Lowe SW. 2009.** The promises and perils of p53. *Nature* **460**: 1085–1086.
- Kubbutat MH, Jones SN, Vousden KH. 1997.** Regulation of p53 stability by Mdm2. *Nature* **387**: 299–303.
- Kulkarni A, Carley H. 2016.** Advances in the recognition and management of hereditary cancer. *British Medical Bulletin* **120**: 123–138.
- Kumaraswamy E, Shiekhattar R. 2007.** Activation of BRCA1/BRCA2-associated helicase BACH1 is required for timely progression through S phase. *Molecular and Cellular Biology* **27**: 6733–6741.
- Kuznetsov SG, Liu P, Sharan SK. 2008.** Mouse ES-cell-based functional assay to evaluate mutations in BRCA2. *Nature medicine* **14**: 875–881.
- Labuschagne CF, Zani F, Vousden KH. 2018.** Control of metabolism by p53 – Cancer and beyond. *Biochimica et Biophysica Acta* **1870**: 32–42.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001.** Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. 2016.** ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* **44**: D862-868.
- Lane DP, Crawford LV. 1979.** T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278**: 261–263.
- Lang GA, Iwakuma T, Suh Y-A, Liu G, Rao VA, Parant JM, Valentin-Vega YA, Terzian T, Caldwell LC, Strong LC, et al. 2004.** Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**: 861–872.
- Lavrut P-M, Le Loarer F, Normand C, Grosos C, Dubois R, Buenerd A, Conter C, Dijoud F, Blay J-Y, Collardeau-Frachon S. 2016.** Small Cell Carcinoma of the Ovary, Hypercalcemic Type: Report of a Bilateral Case in a Teenager Associated with SMARCA4 Germline Mutation. *Pediatric and Developmental Pathology: The Official Journal of the Society for Pediatric Pathology and the Paediatric Pathology Society* **19**: 56–60.
- Lee ASG, Ang P. 2014.** Breast-cancer risk in families with mutations in PALB2. *The New England Journal of Medicine* **371**: 1650–1651.
- Lee KC, Crowe AJ, Barton MC. 1999.** p53-mediated repression of alpha-fetoprotein gene expression by specific DNA binding. *Molecular and Cellular Biology* **19**: 1279–1288.

- Lee M, Daniels MJ, Garnett MJ, Venkitaraman AR. 2011.** A mitotic function for the high-mobility group protein HMG20b regulated by its interaction with the BRC repeats of the BRCA2 tumor suppressor. *Oncogene* **30**: 3360–3369.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O’Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016.** Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291.
- Leu JI-J, Dumont P, Hafey M, Murphy ME, George DL. 2004.** Mitochondrial p53 activates Bak and causes disruption of a Bak-Mcl1 complex. *Nature Cell Biology* **6**: 443–450.
- Levine AJ, Oren M. 2009.** The first 30 years of p53: growing ever more complex. *Nature reviews. Cancer* **9**: 749–758.
- Li FP, Fraumeni JF. 1969.** Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *Journal of the National Cancer Institute* **43**: 1365–1373.
- Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, Miller RW. 1988.** A cancer family syndrome in twenty-four kindreds. *Cancer Research* **48**: 5358–5362.
- Li M-X, Gui H-S, Kwan JSH, Bao S-Y, Sham PC. 2012.** A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Research* **40**: e53.
- Liao Y, Smyth GK, Shi W. 2014.** featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**: 923–930.
- Limacher JM, Frebourg T, Natarajan-Ame S, Bergerat JP. 2001.** Two metachronous tumors in the radiotherapy fields of a patient with Li-Fraumeni syndrome. *International Journal of Cancer* **96**: 238–242.
- Lin D, Chai Y, Izadpanah R, Braun SE, Alt E. 2016.** NPR3 protects cardiomyocytes from apoptosis through inhibition of cytosolic BRCA1 and TNF- α . *Cell Cycle (Georgetown, Tex.)* **15**: 2414–2419.
- Lin T, Chao C, Saito S, Mazur SJ, Murphy ME, Appella E, Xu Y. 2005.** p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nature Cell Biology* **7**: 165–171.
- Lin H-R, Ting NSY, Qin J, Lee W-H. 2003.** M phase-specific phosphorylation of BRCA2 by Polo-like kinase 1 correlates with the dissociation of the BRCA2-P/CAF complex. *The Journal of Biological Chemistry* **278**: 35979–35987.
- Linzer DI, Levine AJ. 1979.** Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17**: 43–52.
- Liu CY, Flesken-Nikitin A, Li S, Zeng Y, Lee WH. 1996.** Inactivation of the mouse Brca1 gene leads to failure in the morphogenesis of the egg cylinder in early postimplantation development. *Genes & Development* **10**: 1835–1843.

- Liu DP, Song H, Xu Y. 2010.** A common gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene* **29**: 949–956.
- Löhr K, Möritz C, Contente A, Dobbstein M. 2003.** p21/CDKN1A mediates negative regulation of transcription by p53. *The Journal of Biological Chemistry* **278**: 32507–32516.
- Lomax ME, Barnes DM, Gilchrist R, Picksley SM, Varley JM, Camplejohn RS. 1997.** Two functional assays employed to detect an unusual mutation in the oligomerisation domain of p53 in a Li-Fraumeni like family. *Oncogene* **14**: 1869–1874.
- Loose M, Malla S, Stout M. 2016.** Real-time selective sequencing using nanopore technology. *Nature Methods* **13**: 751–754.
- Lord CJ, Ashworth A. 2013.** Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nature Medicine* **19**: 1381–1388.
- Lorusso D, Tripodi E, Maltese G, Lepori S, Sabatucci I, Bogani G, Raspagliesi F. 2018.** Spotlight on olaparib in the treatment of BRCA-mutated ovarian cancer: design, development and place in therapy. *Drug Design, Development and Therapy* **12**: 1501–1509.
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.
- Loveday C, Turnbull C, Ruark E, Xicola RMM, Ramsay E, Hughes D, Warren-Perry M, Snape K, Breast Cancer Susceptibility Collaboration (UK), Eccles D, et al. 2012.** Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nature Genetics* **44**: 475–476; author reply 476.
- Lujambio A, Akkari L, Simon J, Grace D, Tschaharganeh DF, Bolden JE, Zhao Z, Thapar V, Joyce JA, Krizhanovsky V, et al. 2013.** Non-cell-autonomous tumor suppression by p53. *Cell* **153**: 449–460.
- Maby P, Tougeron D, Hamieh M, Mlecnik B, Kora H, Bindea G, Angell HK, Fredriksen T, Elie N, Fauquemberg E, et al. 2015.** Correlation between Density of CD8+ T-cell Infiltrate in Microsatellite Unstable Colorectal Cancers and Frameshift Mutations: A Rationale for Personalized Immunotherapy. *Cancer Research* **75**: 3446–3455.
- Mai PL, Best AF, Peters JA, DeCastro RM, Khincha PP, Loud JT, Bremer RC, Rosenberg PS, Savage SA. 2016.** Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. *Cancer* **122**: 3673–3681.
- Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA. 1990.** Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science (New York, N.Y.)* **250**: 1233–1238.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. 2005.** Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* **437**: 376–380.
- Martinez JS, Baldeyron C, Carreira A. 2015.** Molding BRCA2 function through its interacting partners. *Cell Cycle (Georgetown, Tex.)* **14**: 3389–3395.

- Martinez-Useros J, Garcia-Foncillas J. 2016.** The Role of BRCA2 Mutation Status as Diagnostic, Predictive, and Prognosis Biomarker for Pancreatic Cancer. *BioMed Research International* **2016**.
- McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MHN, Eeles RA, Thomas DM, Mitchell G. 2014.** Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nature Reviews. Clinical Oncology* **11**: 260–271.
- McConnell AM, Yao C, Yeckes AR, Wang Y, Selvaggio AS, Tang J, Kirsch DG, Stripp BR. 2016.** p53 Regulates Progenitor Cell Quiescence and Differentiation in the Airway. *Cell Reports* **17**: 2173–2182.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010.** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.
- Mello SS, Attardi LD. 2018.** Deciphering p53 signaling in tumor suppression. *Current Opinion in Cell Biology* **51**: 65–72.
- Mello SS, Valente LJ, Raj N, Seoane JA, Flowers BM, McClendon J, Biegging-Rolett KT, Lee J, Ivanochko D, Kozak MM, et al. 2017.** A p53 Super-tumor Suppressor Reveals a Tumor Suppressive p53-Ptpn14-Yap Axis in Pancreatic Cancer. *Cancer Cell* **32**: 460-473.e6.
- Menon V, Povirk L. 2014.** Involvement of p53 in the repair of DNA double strand breaks: multifaceted Roles of p53 in homologous recombination repair (HRR) and non-homologous end joining (NHEJ). *Sub-Cellular Biochemistry* **85**: 321–336.
- Michl J, Zimmer J, Tarsounas M. 2016.** Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *The EMBO Journal* **35**: 909–923.
- Mihara M, Erster S, Zaika A, Petrenko O, Chittenden T, Pancoska P, Moll UM. 2003.** p53 Has a Direct Apoptogenic Role at the Mitochondria. *Molecular Cell* **11**: 577–590.
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W. 1994.** A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (New York, N.Y.)* **266**: 66–71.
- Millot GA, Carvalho MA, Caputo SM, Vreeswijk MPG, Brown MA, Webb M, Rouleau E, Neuhausen SL, Hansen T v O, Galli A, et al. 2012.** A guide for functional analysis of BRCA1 variants of uncertain significance. *Human Mutation* **33**: 1526–1537.
- Milner J, Medcalf EA, Cook AC. 1991.** Tumor suppressor p53: analysis of wild-type and mutant p53 complexes. *Molecular and Cellular Biology* **11**: 12–19.
- Milner J, Ponder B, Hughes-Davies L, Seltmann M, Kouzarides T. 1997.** Transcriptional activation functions in BRCA2. *Nature* **386**: 772–773.
- Mirkovic N, Marti-Renom MA, Weber BL, Sali A, Monteiro ANA. 2004.** Structure-based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer predisposition. *Cancer Research* **64**: 3790–3797.

- Moll UM, Petrenko O. 2003.** The MDM2-p53 interaction. *Molecular cancer research: MCR* **1**: 1001–1008.
- Momand J, Wu HH, Dasgupta G. 2000.** MDM2--master regulator of the p53 tumor suppressor protein. *Gene* **242**: 15–29.
- Monteiro AN. 2000.** BRCA1: exploring the links to transcription. *Trends in Biochemical Sciences* **25**: 469–474.
- Monteiro AN, August A, Hanafusa H. 1996.** Evidence for a transcriptional activation function of BRCA1 C-terminal region. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 13595–13599.
- Monti P, Ciribilli Y, Jordan J, Menichini P, Umbach DM, Resnick MA, Luzzatto L, Inga A, Fronza G. 2007.** Transcriptional functionality of germ line p53 mutants influences cancer phenotype. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **13**: 3789–3795.
- Moretta J, Berthet P, Bonadona V, Caron O, Cohen-Haguenuer O, Colas C, Corsini C, Cusin V, De Pauw A, Delnatte C, et al. 2018.** [The French Genetic and Cancer Consortium guidelines for multigene panel analysis in hereditary breast and ovarian cancer predisposition]. *Bulletin Du Cancer*.
- Muller PAJ, Vousden KH, Norman JC. 2011.** p53 and its mutants in tumor cell migration and invasion. *The Journal of Cell Biology* **192**: 209–218.
- Nepomuceno TC, De Gregoriis G, de Oliveira FMB, Suarez-Kurtz G, Monteiro AN, Carvalho MA. 2017.** The Role of PALB2 in the DNA Damage Response and Cancer Predisposition. *International Journal of Molecular Sciences* **18**.
- Newman B, Austin MA, Lee M, King MC. 1988.** Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 3044–3048.
- Ng PC, Henikoff S. 2003.** SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**: 3812–3814.
- Nguyen T-AT, Grimm SA, Bushel PR, Li J, Li Y, Bennett BD, Lavender CA, Ward JM, Fargo DC, Anderson CW, et al. 2018.** Revealing a human p53 universe. *Nucleic Acids Research* **46**: 8153–8167.
- Nicholson JK, Lindon JC, Holmes E. 1999.** ‘Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems* **29**: 1181–1189.
- Niculescu AB, Chen X, Smeets M, Hengst L, Prives C, Reed SI. 1998.** Effects of p21(Cip1/Waf1) at both the G1/S and the G2/M cell cycle transitions: pRb is a critical determinant in blocking DNA replication and in preventing endoreduplication. *Molecular and Cellular Biology* **18**: 629–643.

- Nigro JM, Baker SJ, Preisinger AC, Jessup JM, Hostetter R, Cleary K, Bigner SH, Davidson N, Baylin S, Devilee P. 1989.** Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342**: 705–708.
- Nikkilä J, Parplys AC, Pylkäs K, Bose M, Huo Y, Borgmann K, Rapakko K, Nieminen P, Xia B, Pospiech H, et al. 2013.** Heterozygous mutations in PALB2 cause DNA replication and damage response defects. *Nature Communications* **4**: 2578.
- Nyrén P. 2007.** The history of pyrosequencing. *Methods in Molecular Biology (Clifton, N.J.)* **373**: 1–14.
- Olive KP, Tuveson DA, Ruhe ZC, Yin B, Willis NA, Bronson RT, Crowley D, Jacks T. 2004.** Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**: 847–860.
- Oliver SG, Winson MK, Kell DB, Baganz F. 1998.** Systematic functional analysis of the yeast genome. *Trends in Biotechnology* **16**: 373–378.
- Olivos DJ, Mayo LD. 2016.** Emerging Non-Canonical Functions and Regulation by p53: p53 and Stemness. *International Journal of Molecular Sciences* **17**.
- Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M. 2007.** Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Human Mutation* **28**: 622–629.
- Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, Manoukian S, Radice P, van der Luijt RB, van Nesselrooij BPM, et al. 2005.** Classification of BRCA1 missense variants of unknown clinical significance. *Journal of Medical Genetics* **42**: 138–146.
- Pierce LJ, Haffty BG. 2011.** Radiotherapy in the treatment of hereditary breast cancer. *Seminars in Radiation Oncology* **21**: 43–50.
- Pouliot M-C, Kothari C, Joly-Beauparlant C, Labrie Y, Ouellette G, Simard J, Droit A, Durocher F. 2017.** Transcriptional signature of lymphoblastoid cell lines of BRCA1, BRCA2 and non-BRCA1/2 high risk breast cancer families. *Oncotarget* **8**: 78691–78712.
- Pradere B, Lotan Y, Roupret M. 2017.** Lynch syndrome in upper tract urothelial carcinoma: significance, screening, and surveillance. *Current Opinion in Urology* **27**: 48–55.
- Prakash R, Zhang Y, Feng W, Jasin M. 2015.** Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor Perspectives in Biology* **7**: a016600.
- Price KS, Svenson A, King E, Ready K, Lazarin GA. 2018.** Inherited Cancer in the Age of Next-Generation Sequencing. *Biological Research for Nursing* **20**: 192–204.
- Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, Gulden C, Kupfer SS, Swisher EM, Bennett RL, et al. 2012.** ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *The Journal of molecular diagnostics: JMD* **14**: 357–366.

- Qian Y, Chen X. 2008.** ID1, inhibitor of differentiation/DNA binding, is an effector of the p53-dependent DNA damage response pathway. *The Journal of Biological Chemistry* **283**: 22410–22416.
- Rahman N. 2014.** Realizing the promise of cancer predisposition genes. *Nature* **505**: 302–308.
- Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, et al. 2007.** PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics* **39**: 165–167.
- Ransburgh DJR, Chiba N, Ishioka C, Toland AE, Parvin JD. 2010.** Identification of Breast Tumor Mutations in BRCA1 That Abolish Its Function in Homologous DNA Recombination. *Cancer Research* **70**: 988–995.
- Raycroft L, Wu H, Lozano G. 1990.** Transcriptional Activation by Wild-Type But Not Transforming Mutants of the p53 Anti-Oncogene. *Science (New York, N.Y.)* **249**: 1049–1051.
- Reese MG, Eeckman FH, Kulp D, Haussler D. 1997.** Improved splice site detection in Genie. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **4**: 311–323.
- Renaux-Petel M, Charbonnier F, Théry J-C, Fermey P, Lienard G, Bou J, Coutant S, Vezain M, Kasper E, Fourneaux S, et al. 2018.** Contribution of de novo and mosaic TP53 mutations to Li-Fraumeni syndrome. *Journal of Medical Genetics* **55**: 173–180.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015.** Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* **17**: 405–424.
- Robson M, Im S-A, Senkus E, Xu B, Domchek SM, Masuda N, Delaloge S, Li W, Tung N, Armstrong A, et al. 2017.** Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *The New England Journal of Medicine* **377**: 523–533.
- Rodríguez A, D’Andrea A. 2017.** Fanconi anemia pathway. *Current biology: CB* **27**: R986–R988.
- Ros S, Flöter J, Kaymak I, Da Costa C, Houddane A, Dubuis S, Griffiths B, Mitter R, Walz S, Blake S, et al. 2017.** 6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 4 is essential for p53-null cancer cells. *Oncogene* **36**: 3287–3299.
- Rotter V, Friedman H, Katz A, Zerivitz K, Wolf D. 1983.** Variation in antigenic determinants of p53 transformation-related protein obtained from various species. *Journal of Immunology (Baltimore, Md.: 1950)* **131**: 329–333.
- Roy R, Chun J, Powell SN. 2011.** BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews. Cancer* **12**: 68–78.
- Salmon A, Amikam D, Sodha N, Davidson S, Basel-Vanagaite L, Eeles RA, Abeliovich D, Peretz T. 2007.** Rapid development of post-radiotherapy sarcoma and breast cancer in a

patient with a novel germline ‘de-novo’ TP53 mutation. *Clinical Oncology (Royal College of Radiologists (Great Britain))* **19**: 490–493.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. 1977a. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**: 687–695.

Sanger F, Nicklen S, Coulson AR. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463–5467.

Schärer E, Iggo R. 1992. Mammalian p53 can function as a transcription factor in yeast. *Nucleic Acids Research* **20**: 1539–1545.

Scherer SJ, Maier SM, Seifert M, Hanselmann RG, Zang KD, Muller-Hermelink HK, Angel P, Welter C, Scharl M. 2000. p53 and c-Jun functionally synergize in the regulation of the DNA repair gene hMSH2 in response to UV. *The Journal of Biological Chemistry* **275**: 37469–37473.

Schlacher K, Wu H, Jasin M. 2012. A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2. *Cancer cell* **22**: 106–116.

Schwartz R, Schäffer AA. 2017. The evolution of tumour phylogenetics: principles and practice. *Nature reviews. Genetics* **18**: 213–229.

Schwartzberg-Bar-Yoseph F, Armoni M, Karnieli E. 2004. The tumor suppressor p53 down-regulates glucose transporters GLUT1 and GLUT4 gene expression. *Cancer Research* **64**: 2627–2633.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**: 361–362.

Shailani A, Kaur RP, Munshi A. 2018. A comprehensive analysis of BRCA2 gene: focus on mechanistic aspects of its functions, spectrum of deleterious mutations, and therapeutic strategies targeting BRCA2-deficient tumors. *Medical Oncology (Northwood, London, England)* **35**: 18.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* **15**: 7155–7174.

Sharan SK, Morimatsu M, Albrecht U, Lim DS, Regel E, Dinh C, Sands A, Eichele G, Hasty P, Bradley A. 1997. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* **386**: 804–810.

Sharma B, Preet Kaur R, Raut S, Munshi A. 2018. BRCA1 mutation spectrum, functions, and therapeutic strategies: The story so far. *Current Problems in Cancer* **42**: 189–207.

Shaulsky G, Goldfinger N, Ben-Ze’ev A, Rotter V. 1990. Nuclear accumulation of p53 protein is mediated by several nuclear localization signals and plays a role in tumorigenesis. *Molecular and Cellular Biology* **10**: 6565–6577.

- Shaulsky G, Goldfinger N, Rotter V. 1991.** Alterations in tumor development in vivo mediated by expression of wild type or mutant p53 proteins. *Cancer Research* **51**: 5232–5237.
- Shiio Y, Yamamoto T, Yamaguchi N. 1992.** Negative regulation of Rb expression by the p53 gene product. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 5206–5210.
- Shin S, Verma IM. 2003.** BRCA2 cooperates with histone acetyltransferases in androgen receptor-mediated transcription. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 7201–7206.
- Silver DP, Dimitrov SD, Feunteun J, Gelman R, Drapkin R, Lu SD, Shestakova E, Velmurugan S, Denunzio N, Dragomir S, et al. 2007.** Further evidence for BRCA1 communication with the inactive X chromosome. *Cell* **128**: 991–1002.
- Smith EC. 2012.** An overview of hereditary breast and ovarian cancer syndrome. *Journal of Midwifery & Women's Health* **57**: 577–584.
- Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. 2006.** An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human Molecular Genetics* **15**: 2490–2508.
- Sonnenblick A, de Azambuja E, Azim HA, Piccart M. 2015.** An update on PARP inhibitors--moving to the adjuvant setting. *Nature Reviews. Clinical Oncology* **12**: 27–41.
- Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. 2016a.** Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS genetics* **12**: e1005756.
- Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. 2016b.** Correction: Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS genetics* **12**: e1005971.
- Srivastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH. 1990.** Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* **348**: 747–749.
- St Clair S, Giono L, Varmeh-Ziaie S, Resnick-Silverman L, Liu W-J, Padi A, Dastidar J, DaCosta A, Mattia M, Manfredi JJ. 2004.** DNA damage-induced downregulation of Cdc25C is mediated by p53 via two independent mechanisms: one involves direct binding to the cdc25C promoter. *Molecular Cell* **16**: 725–736.
- de Stanchina E, Querido E, Narita M, Davuluri RV, Pandolfi PP, Ferbeyre G, Lowe SW. 2004.** PML is a direct p53 target that modulates p53 effector functions. *Molecular Cell* **13**: 523–535.
- Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S, Shendure J, Parvin JD. 2018.** A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *American Journal of Human Genetics*.

- Starita LM, Machida Y, Sankaran S, Elias JE, Griffin K, Schlegel BP, Gygi SP, Parvin JD. 2004.** BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number. *Molecular and Cellular Biology* **24**: 8457–8466.
- Stehelin D, Varmus HE, Bishop JM, Vogt PK. 1976.** DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**: 170–173.
- Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. 2017.** The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics* **136**: 665–677.
- Stewart MD, Duncan ED, Coronado E, DaRosa PA, Pruneda JN, Brzovic PS, Klevit RE. 2017.** Tuning BRCA1 and BARD1 activity to investigate RING ubiquitin ligase mechanisms. *Protein Science: A Publication of the Protein Society* **26**: 475–483.
- Strong LC, Williams WR, Tainsky MA. 1992.** The Li-Fraumeni syndrome: from clinical epidemiology to molecular genetics. *American Journal of Epidemiology* **135**: 190–199.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH-Y, et al. 2015.** An integrated map of structural variation in 2,504 human genomes. *Nature* **526**: 75–81.
- Sullivan-Reed K, Bolton-Gillespie E, Dasgupta Y, Langer S, Siciliano M, Nieborowska-Skorska M, Hanamshet K, Belyaeva EA, Bernhardt AJ, Lee J, et al. 2018.** Simultaneous Targeting of PARP1 and RAD52 Triggers Dual Synthetic Lethality in BRCA-Deficient Tumor Cells. *Cell reports* **23**: 3127–3136.
- Sumpter R, Sirasanagandla S, Fernández ÁF, Wei Y, Dong X, Franco L, Zou Z, Marchal C, Lee MY, Clapp DW, et al. 2016.** Fanconi Anemia Proteins Function in Mitophagy and Immunity. *Cell* **165**: 867–881.
- Sung P. 2018.** Introduction to the Thematic Minireview Series: DNA double-strand break repair and pathway choice. *The Journal of Biological Chemistry* **293**: 10500–10501.
- Suzuki S, Tanaka T, Poyurovsky MV, Nagano H, Mayama T, Ohkubo S, Lokshin M, Hosokawa H, Nakayama T, Suzuki Y, et al. 2010.** Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 7461–7466.
- Tai YC, Domchek S, Parmigiani G, Chen S. 2007.** Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *Journal of the National Cancer Institute* **99**: 1811–1814.
- Takaoka M, Miki Y. 2018.** BRCA1 gene: function and deficiency. *International Journal of Clinical Oncology* **23**: 36–44.
- Takaoka M, Saito H, Takenaka K, Miki Y, Nakanishi A. 2014.** BRCA2 phosphorylated by PLK1 moves to the midbody to regulate cytokinesis mediated by nonmuscle myosin IIC. *Cancer Research* **74**: 1518–1528.

- Tanaka T, Watanabe M, Yamashita K. 2018.** Potential therapeutic targets of TP53 gene in the context of its classically canonical functions and its latest non-canonical functions in human cancer. *Oncotarget* **9**: 16234–16247.
- Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015.** Sambamba: fast processing of NGS alignment formats. *Bioinformatics (Oxford, England)* **31**: 2032–2034.
- Tazawa H, Tsuchiya N, Izumiya M, Nakagama H. 2007.** Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 15472–15477.
- Thompson D, Easton DF, Goldgar DE. 2003.** A full-likelihood method for the evaluation of causality of sequence variants from family data. *American Journal of Human Genetics* **73**: 652–655.
- Tinat J, Bougeard G, Baert-Desurmont S, Vasseur S, Martin C, Bouvignies E, Caron O, Bressac-de Paillerets B, Berthet P, Dugast C, et al. 2009.** 2009 version of the Chompret criteria for Li Fraumeni syndrome. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **27**: e108-109; author reply e110.
- Toland AE, Andreassen PR. 2017.** DNA repair-related functional assays for the classification of BRCA1 and BRCA2 variants: a critical review and needs assessment. *Journal of Medical Genetics* **54**: 721–731.
- Tomita Y, Marchenko N, Erster S, Nemajerova A, Dehner A, Klein C, Pan H, Kessler H, Pancoska P, Moll UM. 2006.** WT p53, but not tumor-derived mutants, bind to Bcl2 via the DNA binding domain and induce mitochondrial permeabilization. *The Journal of Biological Chemistry* **281**: 8600–8606.
- Tong CWS, Wu M, Cho WCS, To KKW. 2018.** Recent Advances in the Treatment of Breast Cancer. *Frontiers in Oncology* **8**.
- Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J, et al. 2008.** A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Human Mutation* **29**: 1412–1424.
- Tovy A, Spiro A, McCarthy R, Shipony Z, Aylon Y, Allton K, Ainbinder E, Furth N, Tanay A, Barton M, et al. 2017.** p53 is essential for DNA methylation homeostasis in naïve embryonic stem cells, and its loss promotes clonal heterogeneity. *Genes & Development* **31**: 959–972.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010.** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**: 511–515.
- Turk AA, Wisinski KB. 2018.** PARP inhibitors in breast cancer: Bringing synthetic lethality to the bedside. *Cancer* **124**: 2498–2506.

Vaclová T, Gómez-López G, Setién F, Bueno JMG, Macías JA, Barroso A, Urioste M, Esteller M, Benítez J, Osorio A. 2015. DNA repair capacity is impaired in healthy BRCA1 heterozygous mutation carriers. *Breast Cancer Research and Treatment* **152**: 271–282.

Varley JM, Evans DG, Birch JM. 1997. Li-Fraumeni syndrome--a molecular and clinical review. *British Journal of Cancer* **76**: 1–14.

Villani A, Shore A, Wasserman JD, Stephens D, Kim RH, Druker H, Gallinger B, Naumer A, Kohlmann W, Novokmet A, et al. 2016. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. *The Lancet. Oncology* **17**: 1295–1305.

Vogelstein B, Kinzler KW. 2004. Cancer genes and the pathways they control. *Nature Medicine* **10**: 789–799.

Wakasugi M, Kawashima A, Morioka H, Linn S, Sancar A, Mori T, Nikaido O, Matsunaga T. 2002. DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *The Journal of Biological Chemistry* **277**: 1637–1640.

Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King M-C. 2010. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 12629–12633.

Wang EY, Gang H, Aviv Y, Dhingra R, Margulets V, Kirshenbaum LA. 2013. p53 mediates autophagy and cell death by a mechanism contingent on Bnip3. *Hypertension (Dallas, Tex.: 1979)* **62**: 70–77.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics* **10**: 57–63.

Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.

Wheaton K, Muir J, Ma W, Benchimol S. 2010. BTG2 antagonizes Pin1 in response to mitogens and telomere disruption during replicative senescence. *Aging Cell* **9**: 747–760.

Wong L-JC (Ed.). 2013. Next Generation Sequencing: Translation to Clinical Diagnostics. New York: Springer-Verlag.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**: 789–792.

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D. 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science (New York, N.Y.)* **265**: 2088–2090.

Wu GS, Burns TF, McDonald ER, Jiang W, Meng R, Krantz ID, Kao G, Gan DD, Zhou JY, Muschel R, et al. 1997. KILLER/DR5 is a DNA damage-inducible p53-regulated death receptor gene. *Nature Genetics* **17**: 141–143.

- Wu W-S, Heinrichs S, Xu D, Garrison SP, Zambetti GP, Adams JM, Look AT. 2005a.** Slug antagonizes p53-mediated apoptosis of hematopoietic progenitors by repressing puma. *Cell* **123**: 641–653.
- Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian SV, Deffenbaugh A, Goldgar D, Couch FJ. 2005b.** Functional Evaluation and Cancer Risk Assessment of BRCA2 Unclassified Variants. *Cancer Research* **65**: 417–426.
- Xiong Y, Hannon GJ, Zhang H, Casso D, Kobayashi R, Beach D. 1993.** p21 is a universal inhibitor of cyclin kinases. *Nature* **366**: 701–704.
- Yano K, Morotomi K, Saito H, Kato M, Matsuo F, Miki Y. 2000.** Nuclear localization signals of the BRCA2 protein. *Biochemical and Biophysical Research Communications* **270**: 171–175.
- Yeo G, Burge CB. 2004.** Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **11**: 377–394.
- Yu H, Chen JK, Feng S, Dalgarno DC, Brauer AW, Schreiber SL. 1994.** Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**: 933–945.
- Yu X, Wu LC, Bowcock AM, Aronheim A, Baer R. 1998.** The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. *The Journal of Biological Chemistry* **273**: 25388–25392.
- Zerdoumi Y, Aury-Landas J, Bonaïti-Pellié C, Derambure C, Sesboué R, Renaux-Petel M, Frebourg T, Bougeard G, Flaman J-M. 2013.** Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. *Human Mutation* **34**: 453–461.
- Zerdoumi Y, Kasper E, Soubigou F, Adriouch S, Bougeard G, Frebourg T, Flaman J-M. 2015.** A new genotoxicity assay based on p53 target gene induction. *Mutation Research. Genetic Toxicology and Environmental Mutagenesis* **789–790**: 28–35.
- Zerdoumi Y, Lanos R, Raad S, Flaman J-M, Bougeard G, Frebourg T, Tournier I. 2017.** Germline TP53 mutations result into a constitutive defect of p53 DNA binding and transcriptional response to DNA damage. *Human Molecular Genetics* **26**: 2591–2602.
- Zhan Q, Antinore MJ, Wang XW, Carrier F, Smith ML, Harris CC, Fornace AJ. 1999.** Association with Cdc2 and inhibition of Cdc2/Cyclin B1 kinase activity by the p53-regulated protein Gadd45. *Oncogene* **18**: 2892–2900.
- Zhang C, Liu J, Wu R, Liang Y, Lin M, Liu J, Chan CS, Hu W, Feng Z. 2014.** Tumor suppressor p53 negatively regulates glycolysis stimulated by hypoxia through its target RRAD. *Oncotarget* **5**: 5535–5546.
- Zhao YG, Zhao H, Miao L, Wang L, Sun F, Zhang H. 2012.** The p53-induced gene Ei24 is an essential component of the basal autophagy pathway. *The Journal of Biological Chemistry* **287**: 42053–42063.

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM. 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**: 179–184.

Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, Barsyte-Lovejoy D, Al-awar R, Katona BW, Shilatifard A, et al. 2015. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* **525**: 206–211.

Zurer I, Hofseth LJ, Cohen Y, Xu-Welliver M, Hussain SP, Harris CC, Rotter V. 2004. The role of p53 in base excision repair following genotoxic stress. *Carcinogenesis* **25**: 11–19.

Annexe

Article Annexe 1 : Contribution of genotoxic anticancer treatments to the development of multiple primary tumours in the context of germline *TP53* mutations

European Journal of Cancer 101 (2018) 254–262



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com



Original Research

Contribution of genotoxic anticancer treatments to the development of multiple primary tumours in the context of germline *TP53* mutations



Edwige Kasper ^a, Emilie Angot ^b, Elodie Colasse ^b, Lionel Nicol ^c,
Jean-Christophe Sabourin ^b, Sahil Adriouch ^d, Yann Lacoume ^e,
Camille Charbonnier ^a, Sabine Raad ^a, Thierry Frebourg ^{a,*},
Jean-Michel Flaman ^a, Gaëlle Bougeard ^a

^a Normandie Univ, UNIROUEN, Inserm U1245, Rouen University Hospital, Department of Genetics, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

^b Normandie Univ, UNIROUEN, Inserm U1245, Rouen University Hospital, Department of Pathology, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

^c Normandie Univ, UNIROUEN, Inserm U1096, PICTUR – Small Animal Imaging, Rouen University, Rouen, France

^d Normandie Univ, UNIROUEN, Inserm U1234, Rouen University, Rouen, France

^e Animal Facility, Faculty of Medicine and Pharmacy, Rouen University, Rouen, France

Received 5 March 2018; received in revised form 2 May 2018; accepted 8 June 2018

Available online 30 July 2018

KEYWORDS

TP53;
Li-Fraumeni
syndrome;
Multiple primary
cancers;
Anticancer treatments;
Genotoxicity

Abstract Introduction: Li-Fraumeni syndrome (LFS), due to *TP53* germline mutations, is characterised by a remarkably high incidence of multiple primary cancers (MPCs), and the key role of p53 in response to DNA damage questions the contribution of anticancer treatments to MPCs development.

Materials and methods: We first evaluated genotoxicity of X-rays and different classes of conventional chemotherapies, thanks to genotoxicity assays, based on the measurement of transcriptional response to DNA damage and performed in murine splenocytes, either exposed *ex vivo* or extracted from exposed mice. We then exposed a total of 208 *Trp53Δ/Δ*, *wt/Δ* or *wt/wt* mice to clinical doses of X-rays or genotoxic or non-genotoxic chemotherapies. Tumour development was monitored using whole-body magnetic resonance imaging and pathological examination at death.

Results: X-rays and conventional chemotherapies, except mitotic spindle poisons, were found to be genotoxic in both p53 genotoxicity assays. Exposition to X-rays and the topoisomerase inhibitor etoposide, analysed as genotoxic anticancer treatment, drastically increase the

* Corresponding author: Inserm U1245, Faculty of Medicine, 22 boulevard Gambetta, 76183 Rouen, France.
E-mail address: thierry.frebourg@chu-rouen.fr (T. Frebourg).

<https://doi.org/10.1016/j.ejca.2018.06.011>

0959-8049/© 2018 Elsevier Ltd. All rights reserved.

tumour development risk in *Trp53Δ/Δ* and *wt/Δ* mice (hazard ration [HR] = 4.4, 95% confidence interval [CI] [2.2–8.8], $p < 0.001^{***}$ and HR = 4.7, 95% CI [2.4–9.3], $p < 0.001^{***}$, respectively). In contrast, exposure to the non-genotoxic mitotic spindle poison, docetaxel, had no impact on tumour development.

Conclusions: This study shows that radiotherapy and genotoxic chemotherapies significantly increase the risk of tumour development in a LFS mice model. These results strongly support the contribution of genotoxic anticancer treatments to MPC development in LFS patients. Therefore, to reduce the risk of MPCs in germline *TP53* mutation carriers, radiotherapy should be avoided whenever possible, surgical treatment prioritised, and non-genotoxic treatments considered.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Li-Fraumeni syndrome (LFS; MIM#151623), due to germline mutations of the *TP53* tumour suppressor gene, is one of the most severe hereditary predispositions to cancer, characterised by an early age of tumour onset and a broad tumour spectrum [1–5]. The clinical management of LFS patients is particularly challenging, as highlighted by the recently published guidelines for *TP53* mutation carriers [6]. The heterogeneity of the LFS clinical presentation complicates the medical management of mutation carriers. In clinical practice, diagnosis of LFS should be considered in the following clinical situations: a proband with a LFS tumour [premenopausal breast cancer, soft-tissue sarcoma (STS), osteosarcoma, central nervous system tumour and adrenocortical carcinoma (ACC)] before age 46 and at least one first- or second-degree relative with a LFS tumour before age 56; OR a patient with multiple primary tumours, two of which belong to the narrow LFS spectrum and the first being diagnosed before age 46; OR a child with ACC, choroid plexus tumour (CPT), or anaplastic rhabdomyosarcoma (anRMS); OR a female with breast cancer diagnosed before age 31 [7]. An important evolution of the clinical diagnosis LFS criteria is that a familial cancer history is not mandatory to consider the presence of a germline *TP53* mutation. We and other investigators have shown that, independently of the familial history, the *TP53* mutation detection rate in children presenting with ACC can be estimated to 45–66% [7–11], with CPT to 42–100% [7,8], with anRMS up to 73% [12], and in females with breast carcinoma before age 31 to 6% [7,13,14]. Furthermore, we recently reported that *de novo* mutations represent at least 14% of the *TP53* alterations and that approximately 1/5 of these *de novo* mutations occur during embryonic development, resulting in mosaicism [15].

Another remarkable clinical feature of LFS is the very high incidence of secondary tumours [7,16–18]. In the French LFS series, we observed that 43% of the

TP53 mutation carriers develop at least a second primary cancer [7], and this percentage was probably underestimated, as not every patient had benefited from an extensive follow-up. Furthermore, we observed that at least 30% of patients, who had received radiotherapy [7], developed a secondary tumour within the radiation field, confirming other previous observations [19–22]. Finally, we and others described the sequential development of tumours after treatment of a first cancer in *TP53* mutation carriers who were tumour-free until 40 years of age [7,16]. All these clinical observations strongly suggest that chemotherapies and radiotherapy could contribute to the development of secondary tumours in *TP53* mutation carriers, this hypothesis being in full concordance with the key role of p53 in response to DNA damage [23]. Thanks to the development of a new genotoxicity assay based on the measurement of p53 target gene induction, after exposure of human wild-type *TP53* lymphocytes to DNA damaging agents [24], we showed that X-rays and conventional chemotherapies, except mitotic spindle poisons, are strongly genotoxic [24]. We then showed, by measurement of transcriptional response and ChIPSeq experiments, performed in LFS lymphocytes exposed to DNA damaging agents, that germline *TP53* mutations result into a constitutive defect of the binding to DNA and transcriptional response to DNA damage missense mutations with a dominant-negative activity having a more drastic effect than the other types of alterations [25].

The aim of the present study was to explore in a murine model of LFS the contribution of chemotherapies and radiotherapy to the development of secondary tumours in *TP53* mutation carriers.

2. Materials and methods

2.1. Mouse strains and animal studies

Mice were bred, maintained and treated in the Rouen University animal facility. C57Bl/6J mice were obtained

from Charles River Laboratories (Charles River, L'Arbresle, France). B6.129S2-*Trp53^{tm1Tyj}* mice, carrying a large heterozygous deletion encompassing exons 2 to 6 of *Trp53* [26] and subsequently designed *Trp53wt/Δ*, were purchased from Jackson Laboratory (Jackson Laboratory, Bar Harbor, USA). All animal experiments were ethically approved by a certified review board according to French and EU legislations (delivered in October 2015).

2.2. Development of *ex vivo* and *in vivo* p53 genotoxicity assays in mouse

The p53 genotoxicity assay initially developed in human lymphocytes [24] was adapted to mice: splenocytes were purified from adult C57Bl/6J mice, using standard protocols, before treatment for the *ex vivo* genotoxicity assay or after mouse treatment for the *in vivo* genotoxicity assay. Detailed information provided in the Supplemental Material (available online) and Supplementary Table 1.

2.3. Assessment of tumour development in mice

Trp53wt/wt, *Trp53wt/Δ* and *Trp53Δ/Δ* mice were randomised into 12 groups, of at least 15 mice, according to their genotypes treatments: docetaxel 15 mg/kg (Taxotere SANOFI AVENTIS® 80 mg/8 ml), etoposide 15 mg/kg (ETOPOSIDE TEVA® 200 mg/10 ml), whole-body irradiation by 1 Gy of X-rays (Faxitron, Tucson, USA) or NaCl 0.9% as control. Mice were treated at 1.5 month and received treatment thrice, with 4 days of recovery between. Mice were daily monitored with a brief visual examination and were weekly weighed. Tumour development was monthly monitored using magnetic resonance imaging (MRI) after anaesthesia with an intraperitoneal (IP) injection of propofol 200 mg/kg (Propofol MYLAN® 20 mg/ml) and acquisition was achieved with an 1 T Icon (Bruker, Wissembourg, France) dedicated to small animals. Finally, tumour development was assessed by systematic and complete pathological examination at death.

2.4. Statistical analysis

Follow-up was censored at 20 months when all *Trp53Δ/Δ* and *Trp53wt/Δ* mice had died. Kaplan–Meier survival curves and hazard ratios, estimated by a Cox proportional hazard model, were computed to quantify the effect of the different treatments (NaCl, docetaxel, etoposide or X-rays) on cancer mortality, according to the genotypes. Mice that were dead due to causes other than cancer were censored. To estimate the joint effect of each treatment in all LFS model mice (both *Trp53Δ/Δ* and *Trp53wt/Δ*) simultaneously, the Cox regression model was stratified by genotype. Stratification assumed that the same hazard ratios were shared across

genotypes while each genotype group was allowed to follow a separate baseline survival curve. P-values for treatment effects were derived from Wald tests.

3. Results

3.1. X-rays and conventional chemotherapies, except spindle poisons, are genotoxic *ex vivo* and *in vivo* in mice

To evaluate the genotoxicity in mice for anticancer treatments, we developed *ex vivo* and *in vivo* murine versions of the p53 genotoxicity assay (Fig. 1), initially developed in human lymphocytes and based on the measurement by reverse-transcription quantitative multiplex PCR of short fluorescent fragments of the induction of p53 target genes, after DNA damage [24].

We evaluated the genotoxic effect of X-rays and chemotherapies, first, by exposure of primary cultures of wild-type *Trp53* mouse splenocytes (Fig. 1A). The genotoxicity scores of X-rays and chemotherapies, defined as the mean of induction level of six *Trp53* target genes, are indicated in Table 1. In p53 genotoxicity assays, a genotoxicity score equal to 1 indicates the absence of detectable p53 target gene induction and therefore no detectable genotoxicity. X-rays, topoisomerase inhibitors (doxorubicin, epirubicin and etoposide), antimetabolite agents (5FU, gemcitabine), alkylating agents (cisplatin, melphalan) and anti-tumour antibiotic (bleomycin) were found to be genotoxic, as indicated by their high genotoxicity scores (Fig. 1C, Table 1). The drastic reduction of the transcriptional induction of *Trp53* target genes in *Trp53wt/Δ* splenocytes (Table 1) exposed to genotoxic agents demonstrates that this transcriptional response to DNA damage is *Trp53* dependent and is altered in cells harbouring a heterozygous inactivating alteration of *Trp53*.

Western blot analysis of exposed and unexposed cells confirmed that X-rays and genotoxic drugs cause *Trp53* accumulation (Supplementary Fig. 1). In contrast, in this *ex vivo* p53 genotoxicity assay, no genotoxic activity was detected, for mitotic spindle poisons, corresponding either to taxoids (paclitaxel, docetaxel), vinca alkaloids (vindesine, vincristine) or eribulin, as indicated by genotoxicity scores close to 1 (Fig. 1D, Table 1).

We then evaluated *in vivo* genotoxic effects of the treatments by exposing live mice to X-rays or chemotherapies and performing the p53 genotoxicity assay on their splenocytes (Fig. 1B). Doses were calculated to mimic clinical doses administered to cancer patients, on the basis of preclinical studies, taking into account the rapid metabolism in mice and the pharmacokinetics of each molecule (Supplementary Table 2). This *in vivo* version of the p53 genotoxicity assay confirms that X-rays and chemotherapies, except spindle poisons, are genotoxic (Fig. 1E and F, Table 1). Moreover, the *in vivo* assay revealed that cyclophosphamide, an alkylating agent requiring hepatic metabolization, is also genotoxic (Table 1).

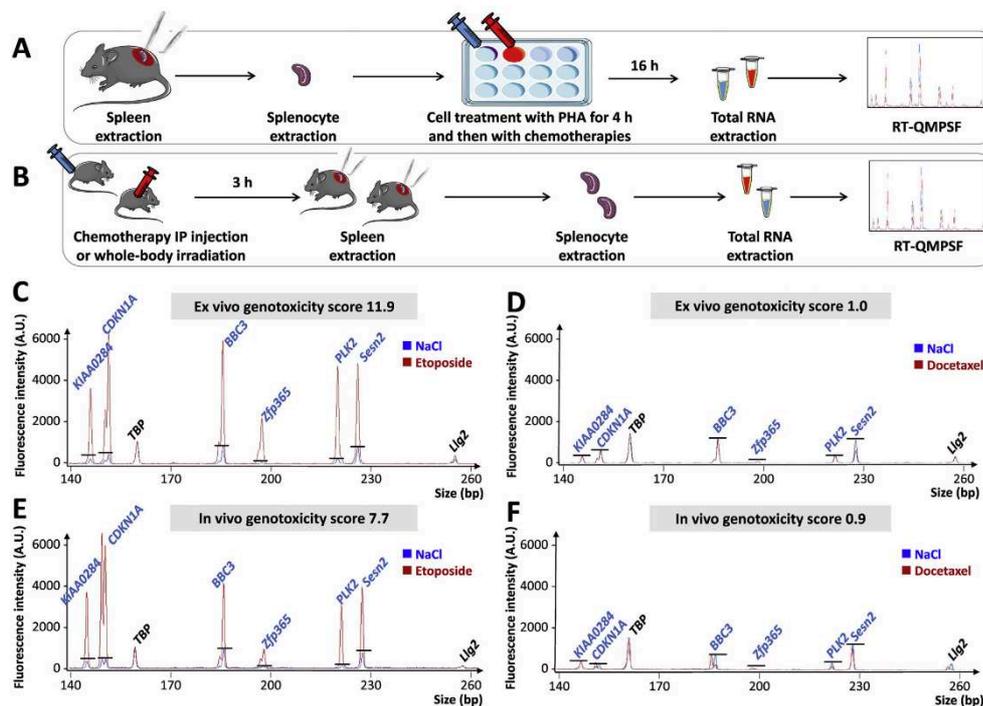


Fig. 1. All conventional chemotherapies except spindle poisons are genotoxic *ex vivo* and *in vivo* in mice. General overview of the *ex vivo* (A) and *in vivo* (B) *Trp53* genotoxicity assays in mice. C–F) RT-QMPSF profiles generated from wild-type *Trp53* murine splenocytes, exposed to etoposide or docetaxel (red profiles) or to NaCl (blue profiles) were superimposed. The different *Trp53* target genes (*KIAA0284*, *CDKN1A*, *BBC3*, *Zfp365*, *PLK2*, *Sesn2*) and the two control genes (*TBP*, *Lig2*) are indicated in blue and black, respectively. The y-axis displays fluorescent intensity in arbitrary units and the x-axis indicates the size of the amplicons in base pairs. For each amplicon, the horizontal bars indicate the height of the peak in unexposed cells. C) and D) *ex vivo* genotoxicity assay performed on wild-type *Trp53* splenocytes exposed to etoposide or docetaxel, respectively. E) and F) *in vivo* genotoxicity assay performed on wild-type *Trp53* mice injected with etoposide or docetaxel, respectively. RT-QMPSF, reverse-transcription quantitative multiplex PCR of short fluorescent fragments; PHA, phytohemagglutinin A; IP, intraperitoneal.

3.2. Increased risk of tumour development in Li-Fraumeni syndrome model mice exposed to X-rays or to genotoxic chemotherapy

We then evaluated the impact of exposure to anticancer treatments on tumour development in *Trp53* Δ/Δ and *Trp53**wt*/ Δ mice. We selected (i) X-rays, (ii) etoposide as genotoxic chemotherapy, as topoisomerase inhibitors were found to be highly genotoxic in p53 genotoxicity assays and (iii) docetaxel as non-genotoxic chemotherapy, as spindle poisons had no detectable genotoxic effect in both assays (Table 1). We used KO *Trp53* mice, harbouring a large deletion of *Trp53*, as LFS mice model, because the common deleterious effect of all germline *TP53* variations is a loss of p53 function [27].

Mice were either whole-body irradiated thrice with 1 Gy (3 Gy in total) or injected thrice with etoposide

(15 mg/kg) or docetaxel (15 mg/kg) and mice treated with NaCl 0.9% were used as negative control. A total of 208 mice assigned into 12 groups, defined by a specific genotype and exposure, were analysed. The primary end-point was death by cancer.

First, we analysed the impact of radiotherapy exposure on *Trp53* Δ/Δ mice, which are known to be highly predisposed to malignancies [26]. Median ages of death by cancer were, in control and irradiated *Trp53* Δ/Δ , 5.8 and 4.5 months, respectively (Fig. 2A). Statistical analysis demonstrated a significant increase of the risk of tumour development in irradiated *Trp53* Δ/Δ mice, as compared with control mice (hazard ratio [HR] = 4.3, 95% confidence interval [CI] [1.7–10.6], $p < 0.001$ ***). In *Trp53**wt*/ Δ mice, mimicking the genetic status of LFS patients, the median age of death by cancer of mice exposed to X-rays was 7.3 months, whereas the median

Table 1
Evaluation of radiotherapy and chemotherapy genotoxicity in mice using the *ex vivo* and *in vivo* Trp53 genotoxicity assays.

Agent	Mechanism of action	<i>Ex vivo</i> Trp53 genotoxicity assay			<i>In vivo</i> Trp53 genotoxicity assay	
		Working condition ^a	Average fold-change of TG expression		Working condition ^a	Average fold-change of TG expression
			Wild-type <i>Trp53</i> splenocytes (genotoxicity score)	Mutant <i>Trp53</i> splenocytes		
NaCl	Negative control	50 µM	1.0 ± 0.0	1.0 ± 0.2	30 mg/kg	1.0 ± 0.0
X-rays	DNA breaks	10 Gy	20.2 ± 4.6	3.9 ± 0.8	5 Gy	25.4 ± 0.7
					1 Gy	6.7 ± 0.4
Doxorubicin	Topoisomerase inhibitor	0.3 µM	19.2 ± 4.8	4.2 ± 1.0	15 mg/kg	3.7 ± 0.3
Epirubicin	Topoisomerase inhibitor	7.5 µM	22.7 ± 4.4	5.8 ± 1.2	30 mg/kg	4.0 ± 1.8
Etoposide	Topoisomerase inhibitor	10 µM	20.7 ± 4.6	4.6 ± 1.0	15 mg/kg	6.5 ± 2.7
Gemcitabine	Antimetabolite agent	10 µM	12.8 ± 1.9	4.2 ± 1.7	20 mg/kg	2.7 ± 0.4
5-FU	Antimetabolite agent	38 µM	13.4 ± 4.3	3.1 ± 1.0	30 mg/kg	2.2 ± 0.8
Cisplatin	Alkylating agent	50 µM	6.3 ± 0.8	2.8 ± 0.8	35 mg/kg	2.6 ± 0.4
Melphalan	Alkylating agent	10 µM	12.4 ± 4.1	2.7 ± 0.5	20 mg/kg	2.1 ± 0.1
Cyclophosphamide	Alkylating agent	Not tested	–	–	300 mg/kg	3.2 ± 0.8
Bleomycin sulphate	Antitumour antibiotic	14 µM	14.1 ± 2.1	3.4 ± 0.3	15 mg/kg	3.7 ± 0.2
Paclitaxel	Spindle poison	7 nM	0.7 ± 0.3	1.2 ± 0.3	5 mg/kg	0.8 ± 0.1
Docetaxel	Spindle poison	3.2 nM	1.1 ± 0.3	1.2 ± 0.2	15 mg/kg	1.0 ± 0.1
Vindesine	Spindle poison	0.3 nM	0.7 ± 0.3	1.2 ± 0.1	5 mg/kg	1.1 ± 0.2
Vincristine	Spindle poison	1 µM	0.6 ± 0.3	1.2 ± 0.3	5 mg/kg	1.2 ± 0.1
Eribulin	Spindle poison	30 nM	0.8 ± 0.3	1.3 ± 0.2	5 mg/kg	1.0 ± 0.1

^a Doses of chemicals were defined on the basis of publications in the field and the Toxnet Chemical Carcinogenesis Research Information System (CCRIS) database (<http://toxnet.nlm.nih.gov/newtoxnet/ccris.htm>).

age of death for control *Trp53wt/Δ* mice was 13.3 months (Fig. 2B). Increased risk of tumour development in irradiated *Trp53wt/Δ* mice was also statistically significant (HR = 4.7, 95% CI [1.5–14.6], $p < 0.01^{**}$). To evaluate the global effect of therapeutic irradiation, we performed statistical analyses stratified by genotype. These analyses showed a highly significant increased risk of tumour development in irradiated LFS model mice (HR = 4.4, 95% CI [2.2–8.8], $p < 0.001^{***}$).

Etoposide-exposed *Trp53Δ/Δ* mice also died prematurely from cancers compared to control mice, with a median age of death by cancer equal to 4.1 months (Fig. 2A). Statistical analysis demonstrated a significant increase risk of tumour development in *Trp53Δ/Δ* mice exposed to etoposide (HR = 4.5, 95% CI [1.9–10.6], $p < 0.001^{***}$), as compared to control mice. In *Trp53wt/Δ* mice exposed to etoposide, the median age of death by cancer was 9.7 months (Fig. 2B) and the increased risk of tumour development was also statistically significant (HR = 6.0, 95% CI [1.8–19.3], $p < 0.01^{**}$). Statistical analyses stratified by genotype also showed a highly significant increased risk of tumour development in LFS model mice exposed to etoposide (HR = 4.7, 95% CI [2.4–9.3], $p < 0.001^{***}$). Early tumour growth in *Trp53wt/Δ* mice exposed to genotoxic treatment could be visualised by sequential MRI, as illustrated in Fig. 3.

In contrast, no increased risk of tumour development was observed in mice exposed to docetaxel, as compared to controls (Fig. 2A and B). The median ages of death by cancer of exposed *Trp53Δ/Δ* and *Trp53wt/Δ* mice were evaluated at 5.2 and 15.9 months, respectively, and

the difference with controls was not statistically different (HR = 1.9, 95% CI [0.8–4.9], $p = 0.143$ for *Trp53Δ/Δ* mice and HR = 0.7, 95% CI [0.2–2.2], $p = 0.506$ for *Trp53wt/Δ* mice). Statistical analyses stratified by genotype confirmed that docetaxel has no impact on the risk of tumour development in LFS model mice (HR = 1.2, 95% CI [0.8–2.5], $p = 0.587$).

Among the *Trp53wt/wt* mice groups, exposed to X-rays, etoposide, docetaxel or NaCl, a single irradiated *Trp53wt/wt* mouse died prematurely from cancer (Fig. 2C). No significant impact on tumour development was observed for *Trp53wt/wt* mice exposed to X-rays (HR = 1.0, 95% CI [0–∞], $p = 1$), to etoposide (HR = 1.0, 95% CI [0–∞], $p = 1$) or to docetaxel (HR = 1.0, 95% CI [0–∞], $p = 1$).

3.3. Genotoxic treatments modify tumour spectrum in *Li-Fraumeni syndrome model mice*

In our study, the majority (86%) of the *Trp53Δ/Δ* control mice developed lymphomas, which is in agreement with the original description of tumour spectrum in this model [26]. In contrast, among *Trp53Δ/Δ* mice exposed to etoposide, only 50% developed lymphomas, 43% developed STS and 7% developed carcinomas (Fig. 4). Similarly, in control *Trp53wt/Δ* mice, the most frequently observed malignancies were lymphomas (42%), whereas in *Trp53wt/Δ* mice exposed to etoposide, we observed 46% of STS, 18% of osteosarcomas, 18% of lymphomas, 9% of carcinomas and 9% of undifferentiated tumours (Fig. 4). X-ray exposure also leads to a

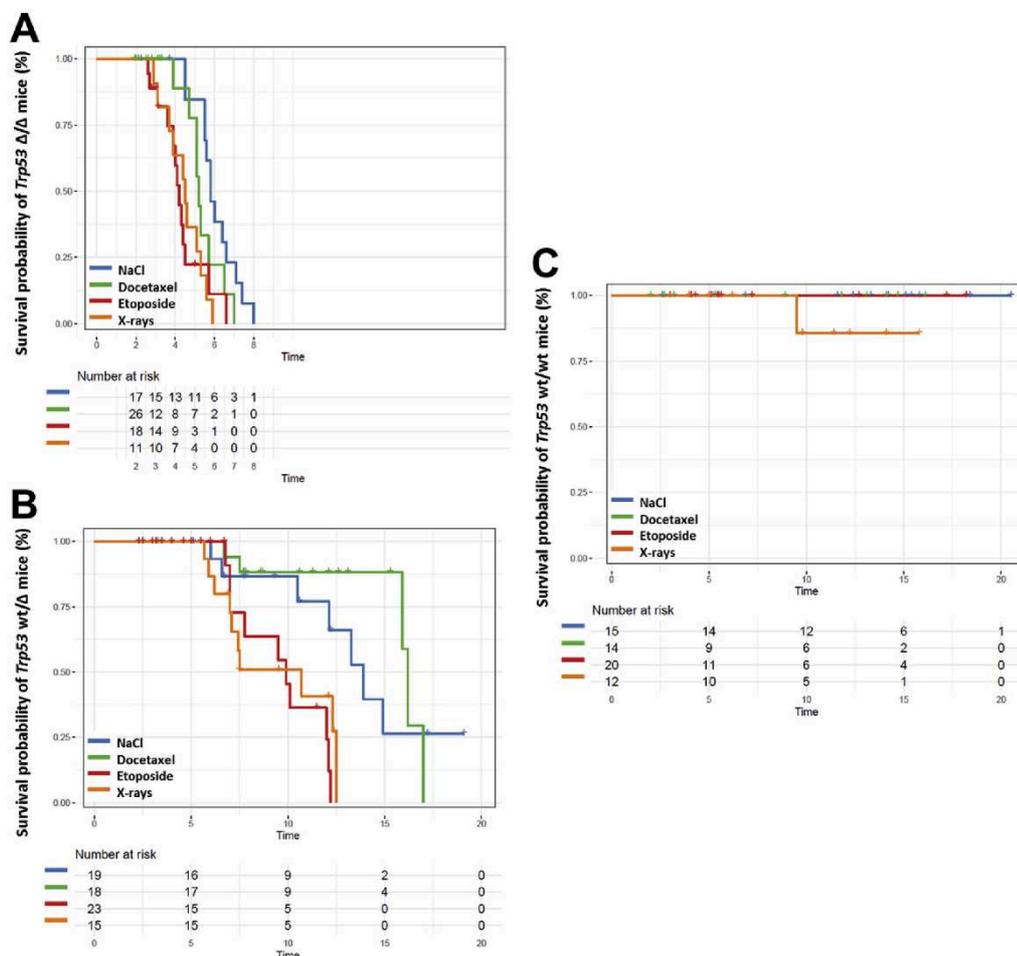


Fig. 2. Survival curves of LFS model mice, dead from cancer, treated with NaCl, docetaxel, etoposide or irradiated. Survival curves of *Trp53* Δ/Δ (A), *Trp53*wt/ Δ (B) and *Trp53*wt/wt (C) mice treated at 1.5 months of age with NaCl (blue), docetaxel (green), etoposide (red) or whole-body irradiated (orange). Censored mice are represented by crosses. LFS, Li-Fraumeni syndrome.

change in tumour spectrum in *Trp53*wt/ Δ mice, with 82% of lymphomas and only 9% of STS (Fig. 4).

4. Discussion

While clinical arguments have strongly suggested that radiotherapy could contribute to the development of secondary tumours in LFS patients [18–22], it has not been experimentally proved yet that anticancer treatments, radiotherapy and chemotherapies, could be responsible for the high incidence of multiple primary tumours observed in *TP53* mutation carriers.

In the first part of this study, we have shown that X-rays and conventional chemotherapies, such as doxorubicin, etoposide, or cisplatin, commonly used in LFS patients are genotoxic *ex vivo* and *in vivo* in mouse, similar to that in human cells. We also demonstrated the genotoxicity of drugs requiring metabolization, such as cyclophosphamide. It should be highlighted that doses of chemotherapies, used in the *ex vivo* p53 genotoxicity assay, correspond to the smallest doses showing sufficient cytotoxic activity on culture cells and that doses used in mice were calculated to mimic clinical doses administered to cancer patients, on the basis of

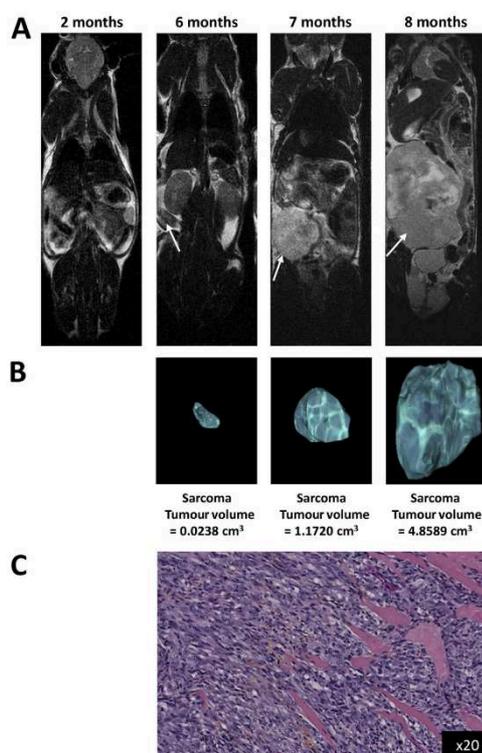


Fig. 3. Whole-body MRI analysis of tumour development in a *Trp53wt/Δ* mouse treated by etoposide. A) The mouse was injected intraperitoneally with etoposide 15 mg/kg at 1.5 months of age and whole-body MRI images (T2-weighted) were acquired at 2, 6, 7 and 8 months. White arrows indicate the detected tumour. B) Three-dimensional representation of tumour development. Tumour volumes were calculated using the free open source OsiriX® imaging software. C) Histopathological analysis after H&E staining of tumour corresponding to an abdominal undifferentiated sarcoma, characterised by spindle and pleomorphic cells infiltrating muscle wall ($\times 20$ magnification). MRI, magnetic resonance imaging.

preclinical studies. The lower values of the genotoxic scores observed in the *in vivo* assay (Table 1) can easily be explained by the pharmacological kinetics of the drugs. The murine p53 genotoxicity assays, such as the assay developed in human lymphocytes [24], did not detect any genotoxicity activity for all tested mitotic spindle poisons, paclitaxel, docetaxel, vindesine, vincristine and eribulin.

In the second part of this work, we used LFS mouse model to study the impact of chemotherapies and X-rays on the kinetics of tumour development. Germline *TP53* mutations can be dichotomised in dominant-negative missense mutations and other types of

alterations [7], but the common deleterious effect of all germline *TP53* variations is the loss of function [27]. Therefore, we considered that *Trp53wt/Δ* mice, harbouring a large deletion of *Trp53*, were a relevant LFS model. The first major result of this study is the significant increased risk, induced by X-rays, of tumour development in LFS mouse models, and in particular, *Trp53wt/Δ* mice (Fig. 2), mimicking the genetic status of LFS patients. The impact of X-rays in *Trp53wt/Δ* mice had already been analysed [28]; however, it is important to highlight that doses used in that study corresponded to five times the doses we used in our study to mimic therapeutic conditions. The second most important result of this study is that a chemotherapy shown to be genotoxic, such as etoposide, also increases the risk of tumour development in LFS model mice (Figs. 2 and 3). We think that we may extrapolate, from these results, that other chemotherapies shown to be genotoxic also increase the risk of tumour development. We provide strong arguments, using docetaxel as a paradigm, that mitotic spindle poisons, in contrast to genotoxic chemotherapies, do not favour tumour development in LFS patients. It should be noticed that exposure to anti-cancer treatment does not increase the risk of tumour development in *Trp53wt/wt* mice (Fig. 2). This observation highlights the fact that the risk of secondary tumours due to genotoxic treatments is particularly increased in LFS model mice. These results were obtained in *Trp53* KO mice, and we might have observed an even greater effect in knock-in mice, harbouring a missense mutation with a dominant-negative effect.

Although the tumour spectrum observed in LFS mouse model is not representative of the tumour spectrum observed in humans, as highlighted by the predominance of lymphomas, our results suggest that radiotherapy and genotoxic chemotherapies might not only increase tumour risk but also modify tumour spectrum (Fig. 4).

In conclusion, our study provides experimental evidence that radiotherapy and genotoxic chemotherapies contribute to MPC development. These results constitute, in addition to clinical observations [7,19–22], strong additional arguments to avoid, whenever possible, radiotherapy and to prioritise surgical treatment in germline *TP53* mutation carriers. In carriers who develop breast carcinomas without lymph node metastasis, complete mastectomy should probably be performed instead of partial mastectomy followed by radiotherapy. Our results suggest that in germline *TP53* mutation carriers with higher stage breast cancer, mitotic spindle poisons, such as taxanes, might constitute a safer chemotherapy, which might reduce the risk of MPCs. The contribution of radiotherapy and chemotherapies to MPCs in LFS constitutes an additional strong argument to perform early detection of tumours using whole-body and breast MRI [6] and to propose prophylactic mastectomy to female *TP53* mutation

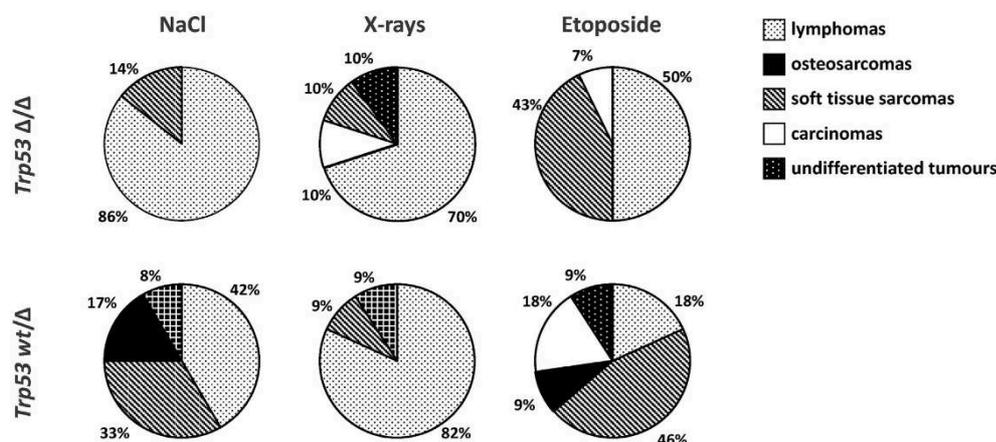


Fig. 4. Tumour spectrum in *Trp53*wt/Δ and *Trp53*Δ/Δ mice exposed to etoposide or X-rays. Pie charts show the relative frequency of tumour types observed according to the treatment.

carriers. Second, medical laboratories in charge of *TP53* testing should ensure fast analysis before treatment and guarantee an expertise in the interpretation of genetic variations, as identification of a germline *TP53* mutation has major therapeutic implications. Finally, our results should prompt the development and evaluation of efficiency and safety in germline *TP53* mutation carriers, of new therapies devoid of genotoxicity, such as immunotherapy, to minimise the risk of MPCs in LFS.

Funding

This work was supported by the ARC Foundation for Cancer Research (Fondation ARC) [grant number POPHARC201401207, 2014], the French National Cancer Institute (INCa), and by the European Regional Development Fund (European Union and Région Normandie). The funding sources had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript and decision to submit the manuscript for publication.

Conflict of interest statement

No conflict of interest to disclose.

Acknowledgements

The authors are grateful to Mario Tosi and Jean-Christophe They for critical review of the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ejca.2018.06.011>.

References

- [1] Li FP, Fraumeni JF. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J Natl Canc Inst* 1969;43:1365–73.
- [2] Li FP, Fraumeni JF, Mulvihill JJ, Blattner WA, Dreyfus MG, Tucker MA, et al. A cancer family syndrome in twenty-four kindreds. *Canc Res* 1988;48:5358–62.
- [3] Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 1990;250:1233–8.
- [4] Srivastava S, Zou ZQ, Pirolo K, Blattner W, Chang EH. Germline transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* 1990;348:747–9. <https://doi.org/10.1038/348747a0>.
- [5] Guha T, Malkin D. Inherited TP53 mutations and the Li-Fraumeni syndrome. *Cold Spring Harb Perspect Med* 2017;7. <https://doi.org/10.1101/cshperspect.a026187>.
- [6] Kratz CP, Achatz MI, Brügières L, Frebourg T, Garber JB, Greer M-LC, et al. Cancer screening recommendations for individuals with Li-Fraumeni syndrome. *Clin Canc Res Off J Am Assoc Canc Res* 2017;23:e38–45. <https://doi.org/10.1158/1078-0432.CCR-17-0408>.
- [7] Bougeard G, Renaux-Petel M, Flaman J-M, Charbonnier C, Ferme P, Belotti M, et al. Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. *J Clin Oncol Off J Am Soc Clin Oncol* 2015;33:2345–52. <https://doi.org/10.1200/JCO.2014.59.5728>.
- [8] Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, et al. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. *J Clin Oncol Off J Am Soc Clin Oncol* 2009;27:1250–6. <https://doi.org/10.1200/JCO.2008.16.6959>.

- [9] Wasserman JD, Novokmet A, Eichler-Jonsson C, Ribcero RC, Rodriguez-Galindo C, Zambetti GP, et al. Prevalence and functional consequence of TP53 mutations in pediatric adrenocortical carcinoma: a children's oncology group study. *J Clin Oncol Off J Am Soc Clin Oncol* 2015;33:602–9. <https://doi.org/10.1200/JCO.2013.52.6863>.
- [10] Wagner J, Portwine C, Rabin K, Leclere JM, Narod SA, Malkin D. High frequency of germline p53 mutations in childhood adrenocortical cancer. *J Natl Canc Inst* 1994;86:1707–10.
- [11] Varley JM, McGown G, Thorncroft M, James LA, Margison GP, Forster G, et al. Are there low-penetrance TP53 Alleles? evidence from childhood adrenocortical tumours. *Am J Hum Genet* 1999;65:995–1006. <https://doi.org/10.1086/302575>.
- [12] Hettmer S, Archer NM, Somers GR, Novokmet A, Wagers AJ, Diller L, et al. Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. *Cancer* 2014;120:1068–75. <https://doi.org/10.1002/cncr.28507>.
- [13] Mouchawar J, Korch C, Byers T, Pitts TM, Li E, McCredie MRE, et al. Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Canc Res* 2010;70:4795–800. <https://doi.org/10.1158/0008-5472.CAN-09-0851>.
- [14] Evans DGR, Moran A, Hartley R, Dawson J, Bulman B, Knox F, et al. Long-term outcomes of breast cancer in women aged 30 years or younger, based on family history, pathology and BRCA1/BRCA2/TP53 status. *Br J Canc* 2010;102:1091–8. <https://doi.org/10.1038/sj.bjc.6605606>.
- [15] Renaux-Petel M, Charbonnier F, Théry J-C, Fermey P, Lienard G, Bou J, et al. Contribution of de novo and mosaic TP53 mutations to Li-Fraumeni syndrome. *J Med Genet* 2017. <https://doi.org/10.1136/jmedgenet-2017-104976>.
- [16] Hisada M, Garber JE, Fung CY, Fraumeni JF, Li FP. Multiple primary cancers in families with Li-Fraumeni syndrome. *J Natl Canc Inst* 1998;90:606–11.
- [17] Izawa N, Matsumoto S, Manabe J, Tanizawa T, Hoshi M, Shigemitsu T, et al. A Japanese patient with Li-Fraumeni syndrome who had nine primary malignancies associated with a germline mutation of the p53 tumour-suppressor gene. *Int J Clin Oncol* 2008;13:78–82. <https://doi.org/10.1007/s10147-007-0692-8>.
- [18] Nutting C, Camplejohn RS, Gilchrist R, Tait D, Blake P, Kneec G, et al. A patient with 17 primary tumours and a germline mutation in TP53: tumour induction by adjuvant therapy? *Clin Oncol R Coll Radiol G B* 2000;12:300–4.
- [19] Heymann S, Delalogue S, Rahal A, Caron O, Frebourg T, Barreau L, et al. Radio-induced malignancies after breast cancer post-operative radiotherapy in patients with Li-Fraumeni syndrome. *Radiat Oncol* 2010;5:104. <https://doi.org/10.1186/1748-717X-5-104>.
- [20] Limacher JM, Frebourg T, Natarajan-Ame S, Bergerat JP. Two metachronous tumours in the radiotherapy fields of a patient with Li-Fraumeni syndrome. *Int J Canc* 2001;96:238–42.
- [21] Salmon A, Amikam D, Sodha N, Davidson S, Basel-Vanagaite L, Eccles RA, et al. Rapid development of post-radiotherapy sarcoma and breast cancer in a patient with a novel germline “de-novo” TP53 mutation. *Clin Oncol R Coll Radiol G B* 2007;19:490–3. <https://doi.org/10.1016/j.clon.2007.05.001>.
- [22] Evans DGR, Birch JM, Ramsden RT, Sharif S, Baser ME. Malignant transformation and new primary tumours after therapeutic radiation for benign disease: substantial risks in certain tumour prone syndromes. *J Med Genet* 2006;43:289–94. <https://doi.org/10.1136/jmg.2005.036319>.
- [23] Kasthuber ER, Lowe SW. Putting p53 in context. *Cell* 2017;170:1062–78. <https://doi.org/10.1016/j.cell.2017.08.028>.
- [24] Zerdoumi Y, Kasper E, Soubigou F, Adriouch S, Bougeard G, Frebourg T, et al. A new genotoxicity assay based on p53 target gene induction. *Mutat Res Genet Toxicol Environ Mutagen* 2015;789–790:28–35. <https://doi.org/10.1016/j.mrgentox.2015.05.010>.
- [25] Zerdoumi Y, Lanos R, Raad S, Flaman J-M, Bougeard G, Frebourg T, et al. Germline TP53 mutations result into a constitutive defect of p53 DNA binding and transcriptional response to DNA damage. *Hum Mol Genet* 2017;26:2591–602. <https://doi.org/10.1093/hmg/ddx106>.
- [26] Jacks T, Remington L, Williams BO, Schmitt EM, Halachmi S, Bronson RT, et al. Tumour spectrum analysis in p53-mutant mice. *Curr Biol* 1994;4:1–7. [https://doi.org/10.1016/S0960-9822\(00\)00002-6](https://doi.org/10.1016/S0960-9822(00)00002-6).
- [27] Freed-Pastor WA, Prives C. Mutant p53: one name, many proteins. *Genes Dev* 2012;26:1268–86. <https://doi.org/10.1101/gad.190678.112>.
- [28] Kemp CJ, Wheldon T, Balmain A. p53-deficient mice are extremely susceptible to radiation-induced tumorigenesis. *Nat Genet* 1994;8:66–9. <https://doi.org/10.1038/ng0994-66>.

Tableau Annexe 1 : Séquences des sondes utilisées en RT-MLPA

Gène	Séquences des sondes	Taille (pb)
KIAA0284	G- 5'-GTGCCAGCAAGATCCAATCTAGAGGGTGCAGAAACAGCTGGAAG-3' D- 5'-/5Phos/TTATCAATGCCATCGTGGACCCCACTCCAACCCTTAGGGAACCC-3'	88
PODXL	G- 5'-GTGCCAGCAAGATCCAATCTAGACTCTCCCAGAGGAAGGACCAG-3' D- 5'-/5Phos/CAGCGGCTAACAGAGGAGCTCCAACCCTTAGGGAACCC-3	82
RRAD	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCGCCTGAAGCAGAGGCAGCAG-3' D- 5'-/5Phos/GGCACACCTATGATCGCTCCATTGCATCTCCAACCCTTAGGGAACCC-3'	94
CABYR	G- 5'-GTGCCAGCAAGATCCAATCTAGACAAACGTCGCAAAGCAGAACTGAAAAGT-3' D- 5'-/5Phos/ATCCAGAAATGACGCTGTCTGGGCATCATCATCATTCCTCCAACCCTTAGGGAACCC-3'	109
TP53I3	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCATCCGCCTTCCAGCTGTTACATCTTGTGG-3' D- 5'-/5Phos/GAAATGTTTCAGGCTGGAGACTATGTGCTAATCATCATCATTCCTCCAACCCTTAGGGAACCC-3'	118
GLS2	G- 5'-GTGCCAGCAAGATCCAATCTAGAGCAAAGTGAATCCTTTTGCCAAGGACAG-3' D- 5'-/5Phos/GTGGGGCAACATTCCCCTGGCATCATCATCATTCCTCCAACCCTTAGGGAACCC-3'	103

SULF2	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCAAGTGGACGACTCCATGGAGACG-3'	105
	D- 5'-/5Phos/ATTTACAACATGCTGGTTGAGACGGGCATCATCATCCAACCCTTAGGGAACCC-3'	
EPS8L2	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCCCTCGCAGTACCACGTCCAG-3'	100
	D- 5'-/5Phos/CACCTGGCCACATTCATCATGGAATCATCATCATCCAACCCTTAGGGAACCC-3'	
SESN1	G-5'- GTGCCAGCAAGATCCAATCTAGACATCATCAGATGCCGGTCAACTCAGCAGAAAATGTTTCT-3'	124
	D- 5'-/5Phos/GTAAGTGATTCTTTCTTTGAGGTTGAAGCCCTCATGCATCATCTCCAACCCTTAGGGAACCC-3'	
FHL2	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCATCATCCAGGAATGCAAGAAGACCATCATGCCAG-3'	127
	D- 5'-/5Phos/GTACCCGCAAGATGGAGTACAAGGGCATCATCATCATCATCTCCAACCCTTAGGGAACCC -3'	
TP53 Exon 4-5	G- 5'-GTGCCAGCAAGATCCAATCTAGAGAGCGCTGCTCAGATAGCGATG-3'	85
	D- 5'-/5Phos/GTCTGGCCCCCTCCTCAGCATCTCCAACCCTTAGGGAACCC-3'	
TP53 Exon 6-7	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCTCACCATCATCACACTGGAAGACTCCAG-3'	115
	D- 5'-/5Phos/TGGTAATCTACTGGGACGGAACAGCTTTATCATCATCTCCAACCCTTAGGGAACCC-3'	
TBP	G- 5'-GTGCCAGCAAGATCCAATCTAGACATCATCCCCGAAACGCCGAATATAATCCCAAG-3'	112
	D- 5'-/5Phos/CGGTTTGCTGCGGTAATCATGAGGATAAGCATCATCATCCAACCCTTAGGGAACCC-3'	

RIC8B	G- 5' -GTGCCAGCAAGATCCAATCTAGACATCATGACGGTAGACAGTTGGAAGGTGCATAAAGAG-3'	120
	D- 5' -/5Phos/AGTGATTCTCATCAGTTCCGTGTAATGGCAGCATCATCATCTCCAACCCTTAGGGAACCC-3'	
<hr/>		
MPP5	G- 5' -GTGCCAGCAAGATCCAATCTAGACATCATCGCTTTGCATCTGCAGTTCCTC-3'	97
	D- 5' -/5Phos/ATACAACCCGGAGTAGGCGAGACATCATCCAACCCTTAGGGAACCC-3'	
<hr/>		
Primers Universels	F-6 FAM-5'-GTGCCAGCAAGATCCAATCTAGA-3'	
	R-5'-GGGTTCCCTAAGGGTTGGA-3'	

G : sonde Gauche ; D sonde Droite, F : Forward ; R : Reverse ; Phos : Phosphorylation ; 6Fam : Florescent 6-Fam-labeled.

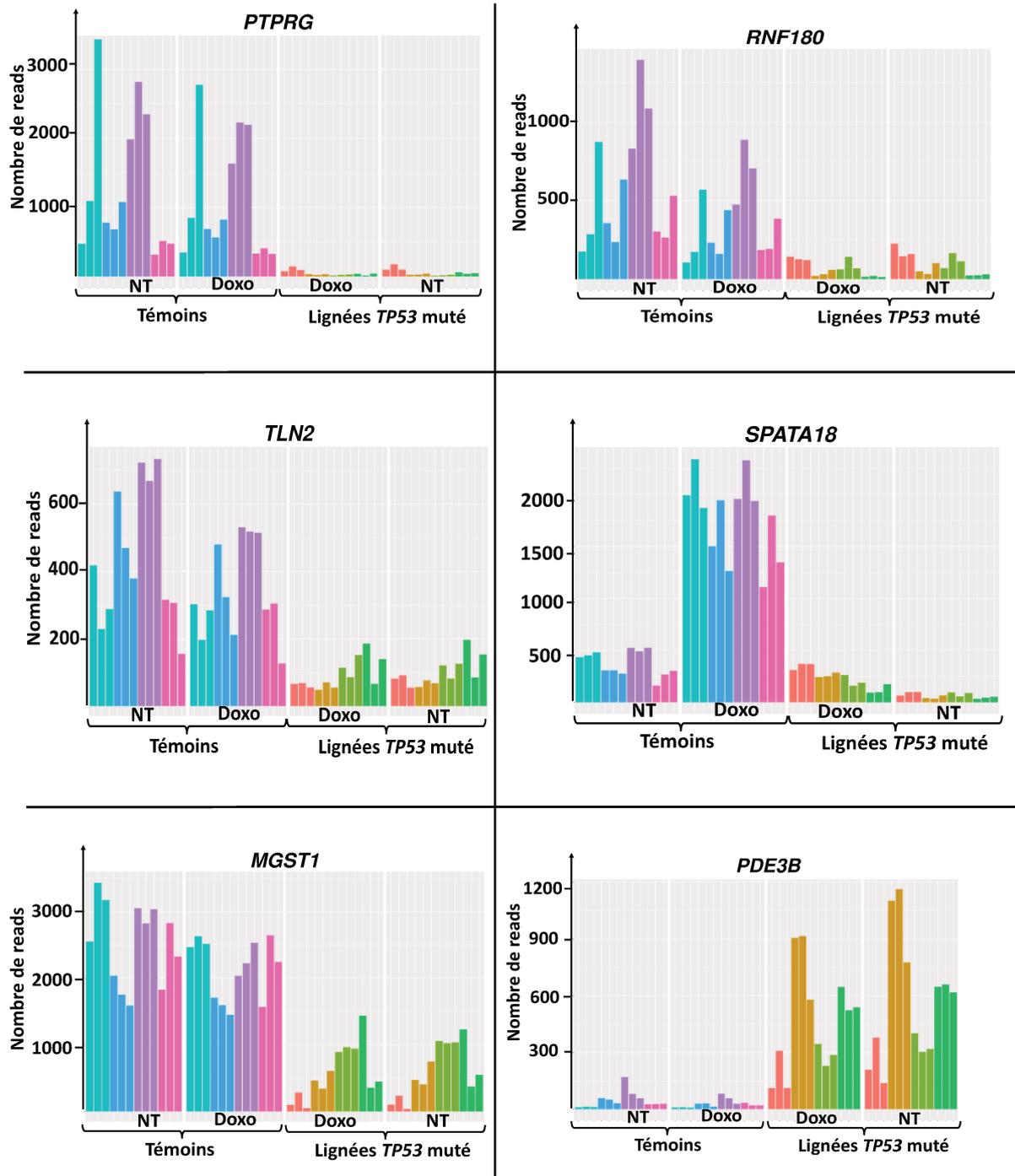
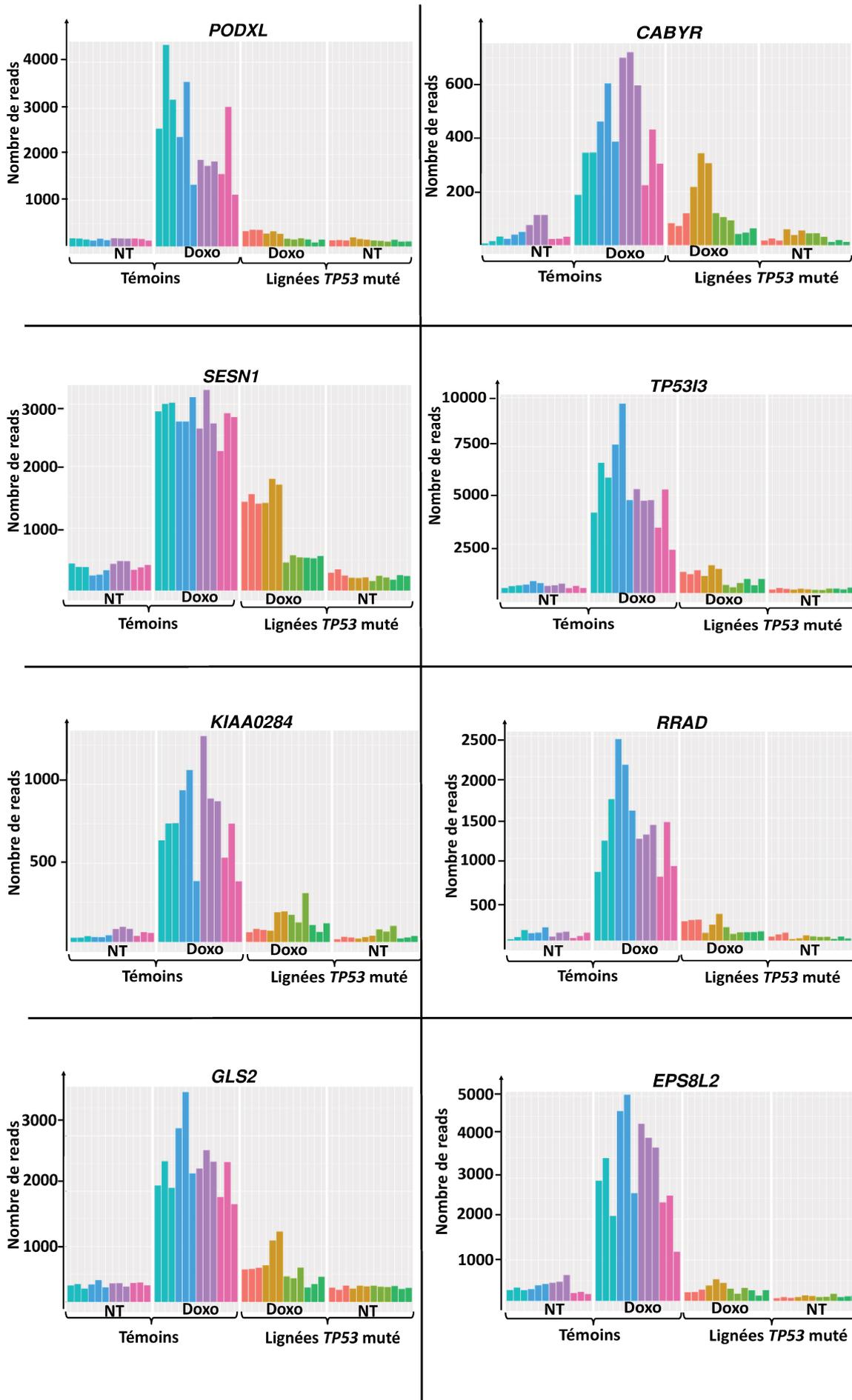


Figure Annexe 1 : Nombre de reads obtenus pour les 6 gènes en RT-QMPSF à l'issu de l'analyse de RNA-Seq chez les témoins et chez les mutants *TP53* sans traitement (NT) ou avec traitement à la doxorubicine (Doxo)



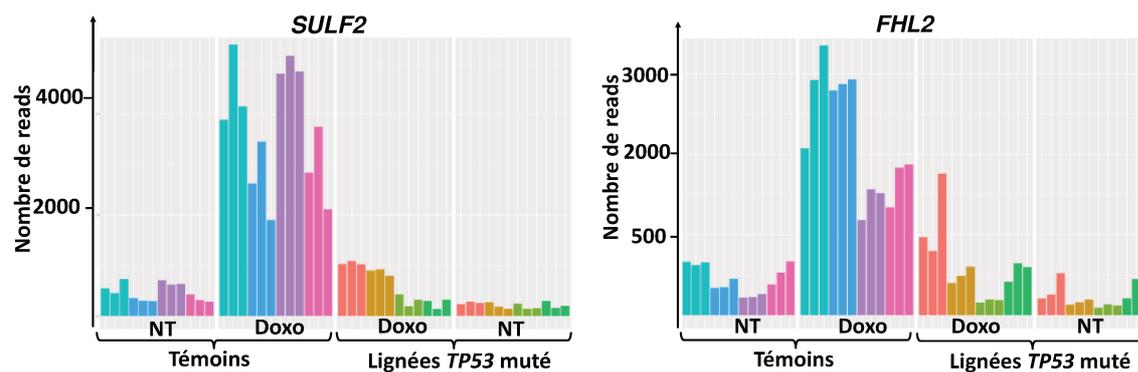


Figure Annexe 2 : Nombre de reads obtenus pour les 10 gènes en RT-MLPA à l'issu de l'analyse de RNA-Seq chez les témoins et chez les mutants *TP53* sans traitement (NT) ou avec traitement à la doxorubicine (Doxo)

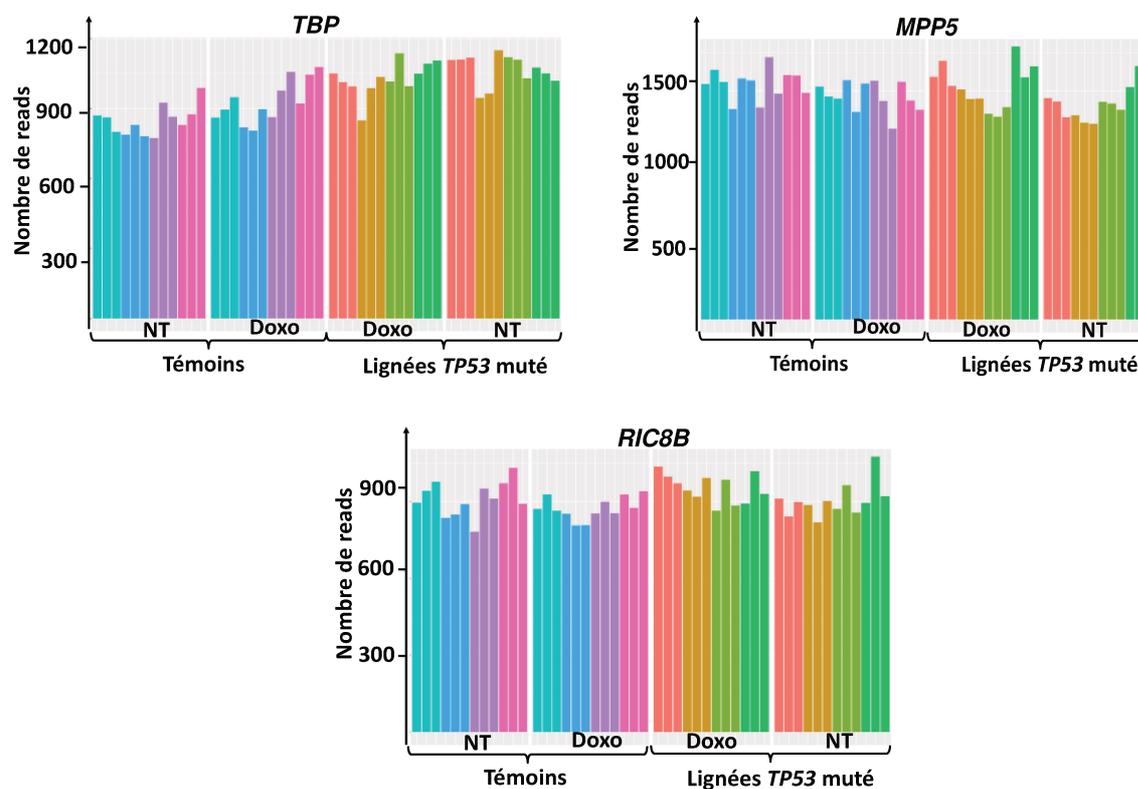


Figure Annexe 3 : Nombre de reads obtenus pour les 3 gènes contrôles en RT-MLPA à l'issu de l'analyse de RNA-Seq chez les témoins et chez les mutants *TP53* sans traitement (NT) ou avec traitement à la doxorubicine (Doxo).

RESUME

L'identification des mutations constitutionnelles à l'origine d'une prédisposition génétique au cancer est essentielle à la prise en charge médicale des patients et de leurs familles. Depuis l'implémentation des technologies de séquençage à haut-débit dans les laboratoires diagnostiques, le principal défi n'est plus la détection des variations génétiques mais leur interprétation et leur classification. La question de l'interprétation de la variation est particulièrement cruciale lorsqu'elle conditionne la stratégie thérapeutique. Ainsi, il est essentiel de disposer de tests simples adaptables en routine diagnostique pour faciliter l'interprétation des variations génétiques. Dans ce contexte, nous avons utilisé un test fonctionnel développé par notre équipe pour classer des variations dans le gène *TP53* à l'origine du syndrome de Li-Fraumeni et pour appréhender la corrélation génotype - phénotype chez les patients LFS. Dans un deuxième temps, nous avons évalué la pertinence d'une approche multi-omique (RNA-Seq et métabolomique) pour discriminer les cellules sauvages des cellules avec mutation hétérozygote du gène *TP53* ou des gènes *BRCA* impliqués dans la prédisposition génétique aux cancers du sein et de l'ovaire. Sur la base des données de transcriptome, un modèle mathématique a été développé pour détecter les variants correspondant à des mutations délétères. Nous avons ensuite sélectionné les biomarqueurs les plus discriminants pour les intégrer dans un test fonctionnel de RT-MLPA dédié à la voie p53. Nous avons enfin adapté cet essai pour qu'il soit réalisable sur une simple prise de sang, sans immortalisation des lymphocytes du patient.

Mots clés : Prédispositions génétiques au cancer, Approche multi-omique, Métabolomique, RNA-seq, Tests fonctionnels, *TP53*, *BRCA*

SUMMARY

The identification of the constitutional mutation responsible for a genetic predisposition to cancer is essential to the clinical management of the patient and its relatives. With the implementation of high-throughput sequencing to the diagnostic routine of these pathologies, the challenge no longer lies within the detection of alterations but in their biological and clinical interpretation. While specific treatments are emerging, simple functional assays to help with the interpretation of the detected variants are needed. In this context, we used a functional test developed by our team to classify variations in the *TP53* gene responsible for Li-Fraumeni syndrome and to understand the genotype-phenotype correlation in LFS patients. On the other hand, we assessed the relevance of a multi-omic approach (RNA-Seq and metabolomics) to discriminate wild-type cells from cells with a deleterious heterozygous mutation in *TP53* or in the *BRCA* genes implicated in genetic predisposition to breast and ovarian cancers. Based on the transcriptomic data, a mathematical model has been developed to detect variants corresponding to deleterious mutations. Then we selected the most discriminating biomarkers and integrated them into a RT-MLPA functional assay dedicated to the p53 pathway. We finally adapted this test to be feasible on a simple blood test, without immortalization of the patient's lymphocytes.

Keywords : Genetic cancer predispositions, Multi-omic approach, RNA-Seq, Metabolomics, Functional assays, *TP53*, *BRCA*