



Détection et contribution de variants rares constitutionnels dans les formes précoces de cancer du sein : Apports du Séquençage de Nouvelle Génération.

Jean-Christophe Thery

► To cite this version:

Jean-Christophe Thery. Détection et contribution de variants rares constitutionnels dans les formes précoces de cancer du sein : Apports du Séquençage de Nouvelle Génération.. Médecine humaine et pathologie. Normandie Université, 2019. Français. NNT : 2019NORMR111 . tel-02501260

HAL Id: tel-02501260

<https://theses.hal.science/tel-02501260>

Submitted on 6 Mar 2020

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 MOLECULAIRES ET CELLULAIRES DE LA BIOLOGIE

Préparée au sein de l'Université de Rouen Normandie

Détection et Contribution de Variants Rares Constitutionnels dans les Formes Précoces de Cancer du Sein : Apports du Séquençage de Nouvelle Génération

**Présentée et soutenue par
Jean-Christophe THERY**

**Thèse soutenue publiquement le 19 Décembre 2019
devant le jury composé de**

Mme Marie-Pierre BUISINE	PUPH, CHRU de Lille	Rapporteur
Mme Frédérique PENAULT-LLORCA	PUPH, Centre Jean Perrin, Clermont-Ferrand	Rapporteur
Mme Brigitte BRESSAC	Pharmacienne spécialiste de Centre, Institut Gustave Roussy, Villejuif	Examineur
Mr Frédéric DI FIORE	PUPH, CHRU et Centre H. Becquerel de Rouen	Examineur
Mr Thierry FREBOURG	PUPH, CHRU de Rouen	Directeur de thèse

Thèse dirigée par Thierry Frébourg et Isabelle Tournier, Inserm 1245, Rouen



Résumé

Si le syndrome de prédisposition héréditaire au cancer du sein et de l'ovaire constitue une entité reconnue et supportée par l'identification de variations délétères sur les gènes *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* et *RAD51D*, et si le cancer du sein de la femme jeune (avant 31 ans) est intégré dans le spectre du syndrome de Li-Fraumeni lié aux altérations de *TP53*, une large fraction des patientes adressées en consultation d'oncogénétique pour ce motif demeure orpheline de diagnostic moléculaire. La connaissance du génome humain et l'avènement du séquençage de nouvelle génération ont permis des avancées considérables, notamment dans l'observation de la très grande variabilité du génome et de la survenue de variations *de novo*.

Nous avons ainsi appliqué ces outils et ces concepts au cancer du sein de la femme jeune, afin de tenter d'identifier de nouveaux déterminants moléculaires constitutionnels. **Dans une première approche basée sur la réalisation d'exomes soustractifs** pour des trios parents - enfant, nous avons recherché des variants *de novo* délétères et effectivement identifié un variant rare *de novo* et délétère sur le gène *INHBA* dans le contexte d'un cancer de l'ovaire chez une jeune femme. Cette approche n'a cependant pu être reproduite dans le contexte d'un cancer du sein précoce. Nous avons également tenté une approche par exomes comparatifs dans une famille remarquable avec survenue de cancers du sein précoces sur trois générations, sans variation délétère identifiée commune à ces individus. **Dans une seconde approche basée sur un panel de 201 gènes impliqués dans la cancérogenèse**, nous avons tenté d'identifier des variants délétères ou des enrichissements en variants rares dans une cohorte de cancers du sein précoces. Nous avons identifié une variation en mosaïque de *TP53*, sans autre détection de variations formellement délétères parmi 30 patientes atteintes de cancers du sein avant 31 ans. Un enrichissement non significatif en variants rares affectant les voies de la réparation de l'ADN a été néanmoins mis en évidence, suggérant des études plus larges ciblant ces voies. **Enfin, nous avons recherché spécifiquement des variants *de novo* en mosaïque de *TP53*** dans le contexte du cancer du sein de la femme jeune ou de cancers pédiatriques, et démontré ainsi la prévalence relativement importante de ces événements. Ces observations supportent la nécessité d'user d'un séquençage de forte profondeur et de ne pas restreindre les indications d'analyses de *TP53* aux seules situations familiales évocatrices.

Mots-clés : Cancer du sein précoce - Séquençage de nouvelle génération - Séquençage complet d'exome - Panel de gènes - Variants *de novo* - Variants en mosaïques

Abstract

Despite previous identifications of deleterious variants on *BRCA1*, *BRCA2*, *PALB2*, *RAD51C* and *RAD51D* supporting the hereditary breast and ovarian cancer syndrom, and the contribution of *TP53* mutations in very early-onset breast carcinomas, a large fraction of patients suggestive of Medelian disease remains without molecular diagnosis. In the past years, sequencing of the Human genome and next-generation sequencing offered major advances, in particular in the field of genome variability and *de novo* variants.

We applied these new tools and concepts in the context of very early-onset breast carcinomas, in order to identify new molecular germline determinants. First, we dealt with **soustractive exomes**, in parents - child trios, and succeed in the identification of a deleterious *de novo* variant in the *INHBA* gene, in the context of very early-onset of ovarian cancer. However, we have failed with this approach in a second trio with an index affected by early-onset breast carcinoma. We also tried a **comparative exome sequencing approach** in a remarkable pedigree with multiple probands affected by early-onset breast carcinomas, without identification of a shared deleterious variant. Secondly, we used a home-made **201 genes panel** assuming that genes somatically affected in cancers might be altered in inherited conditions. We analyzed a cohort of very early-onset breast carcinomas, and identified a mosaic *TP53* variation. Moreover, we identified some interesting candidate variants and observed a non-significant trend of rare variants enrichment in the DNA repair pathway. Finally, we designed a specific *TP53* gene capture in order **to detect mosaic variants** in pediatric cancers and very early-onset breast carcinomas. We confirmed the clinically significant prevalence of these alterations, which support *TP53* analysis in these conditions even in sporadic presentations.

Keywords : Very early-onset breast carcinoma - Next-generation sequencing - Whole Exome Sequencing - Genes panel - *De novo* variants - Mosaic variants

Liste des Abréviations

Par ordre alphabétique :

AB	Balance Allélique
ACMG/AMP	American College of Medical Genetics and Genomics / Association of Molecular Pathology
ASCO	American Society of Clinical Oncology
BI-RADS	Breast Imaging Reporting And Data System
CDS	Coding DNA Sequence
CNV	Copy Number Variation
ESE	Exonic Splicing Enhancer
ESP	Exome Sequencing Projet
ESS	Exonic Splicing Silencer
FREX	France Exome Project
Gnomad	Genome Aggregation Database
HBOC	Hereditary Breast and Ovarian Cancer
HDGC	Hereditary Diffuse Gastric Cancer syndrome
HGMD	Human Gene Mutation Database
HLBC	Hereditary Lobulaire Breast Cancer
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
ICL	Interstrand CrossLinking
IGCLC	International Gastric Cancer Linkage Consortium
LFS	Li-Fraumeni Syndrom
LOEUF	Loss-of-function Observed Expected Upper bound Fraction
MAF	Minor Allelic Frequency
MSI	Microsatellite Instability
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
OR	<i>Odds Ratio</i>
PAF	Polypose Adénomateuse Familial
SERM	Selective Estrogen Receptor Modulator
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
VSI	Variation de Signification Inconnue
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Remerciements

Je remercie sincèrement les membres du Jury de me faire l'honneur d'expertiser cette Thèse et d'apprécier ma contribution : Madame Frédérique Penault-Llorca, Madame Marie-Pierre Buisine, Madame Brigitte Bressac, Monsieur Frédéric Di Fiore.

Dans chaque réalisation personnelle se cache une oeuvre collective, et je reprendrais volontiers ici les mots d'Antoine de Saint Exupéry : "La pierre n'a point d'espoir que d'être autre chose que pierre, mais de collaborer, elle s'assemble et devient temple".

Ces années dans l'unité Inserm 1245 et cette Thèse de Science marquent mon intérêt pour la Recherche fondamentale et appliquée dans le domaine de l'oncogénétique, domaine découvert quelques années plus tôt dans le contexte de mon Master 2. Je souhaiterais en préambule de ce travail remercier chaleureusement toutes les personnes dont l'intellect, la bienveillance et l'amitié m'ont permis d'évoluer, de m'enrichir et de me réaliser dans cette voie de la Recherche au cours de ces années.

Au sein de l'unité Inserm, je citerai en premier lieu le Professeur Thierry Frébourg pour m'avoir accueilli et gratifié de sa confiance en acceptant de diriger cette Thèse. Vous êtes pour plusieurs générations de médecins et d'étudiants l'incarnation de la Recherche, dans la passion, l'excitation intellectuelle, la rigueur scientifique mais surtout dans l'exigence de la translation de cette recherche au service des patients et des familles. Ici se trouve en effet la plus grande source d'énergie, celle qui donne toute cohérence et toute finalité à ces temps passés à la réflexion scientifique. Je vous exprime ici toute ma gratitude et espère avoir été la hauteur de vos enseignements. Je suis également infiniment reconnaissant à Isabelle Tournier pour avoir co-dirigé cette Thèse et m'avoir fait pénétrer dans le monde du séquençage haut débit. Tu as eu la patience et la pédagogie de m'initier aux manipulations, nous avons partagé les moments d'enthousiasme et parfois les moments de déception. Une amitié est née de ces années, tu as été pour moi une belle rencontre. Je garde en mémoire aussi Audrey Killian, toute première personne à m'avoir initié au fonctionnement du Laboratoire lors de mon Master 2. Je remercie tous les membres de l'unité qui m'ont accueilli, guidé, encouragé, en particulier Myriam, Sophie, Raphaël et Olivier pour la bioinformatique, Pierre et Etienne avec qui j'ai partagé des moments de réflexion et des manips, Laurent, Dominique et Sophie de Caen, Pascaline, Gaëlle, Stéphanie et Edwige avec qui nos échanges ont été particulièrement stimulants. Françoise Charbonnier, tu occupes une place prépondérante dans ces remerciements pour l'aide apportée et l'amitié que nous avons développées. Mémoire du Laboratoire, tu te nourris de ces générations d'étudiants que tu accompagnes, guides et encourages dans leurs parcours. Tu m'as particulièrement aidé pour les dernières manipulations et a toujours été réceptive à mes demandes. Je remercie également Emilie Bouvignies pour m'avoir aidé à désarchiver et à doser certains ADN, Nathalie Porchet pour m'avoir aidé dans les aspects administratifs de ce parcours, Rosemonde sans qui le Laboratoire ne pourrait constituer un environnement de travail. Je remercie les différentes instances ayant soutenu financièrement l'unité Inserm et particulièrement ces travaux autour du cancer du sein précoces, notamment Rose sur Green et la Ligue contre le Cancer.

Au sein du Centre Henri Becquerel, je citerai en premier lieu Emmanuel Blot. Tu as été mon premier mentor, celui qui a contribué à ma vocation d'oncologue médical et m'a initié à la Recherche. Je remercie Corinne Veyret pour m'avoir permis de bénéficier d'un environnement aménagé durant mes deux premières années de Thèse, Frédéric pour m'avoir encouragé dans la finalisation de ce travail. Je remercie l'ensemble de mes collègues oncologues pour leur amitié et leurs encouragements, pour m'avoir suppléé durant ces moments de présence au Laboratoire : Cécile, Florian, Olivier, Marianne, Sophie, Isabelle, Cristina, Cristian, Camille et Maxime, vous êtes mes collègues et amis, nous partageons des prises en charge difficiles que je ne pourrais affronter sans votre présence à mes côtés. Mes deux secrétaires successives, Isabelle et Catherine, merci d'avoir respecté ces temps au Laboratoire et de m'avoir aidé à organiser ce planning partagé.

Nous ne pouvons être efficient sur le plan professionnel sans être épanoui et nourri sur le plan personnel et familial. Merci à mes parents pour leur soutien indéfectible, pour m'avoir transmis l'amour du prochain et la passion de la Médecine. Merci à mes soeurs et belle-soeur, à mes beaux-parents, à mes amis, de contribuer à mon équilibre personnel. Et merci à toi, Cathy, aimée et aimante depuis nos 15 ans, Simon, Sarah, Sasha et Yaël, nos quatre plus belles réalisations, d'être mes raisons d'être.

Enfin, je dédicace ce travail à tous mes patient(e)s, dont la souffrance, les drames et parfois les victoires constituent mes sources de motivation. Que dire à une jeune femme affectée à 25 ans d'un cancer du sein triple négatif, que dire à ses parents lorsque, disparue, balayée par la maladie métastatique, nous ne pouvons même pas apporter l'explication de l'origine de cette injustice. Beaucoup des patients étudiés ici sont certainement disparus, qu'ils sachent que nul ne peut les oublier et que nous n'aurons de cesse de tenter de percer les mystères de cette affection afin d'améliorer les prises en charge et peut-être un jour, être en mesure de mieux prévenir, mieux guérir, mieux soigner. Françoise, mon amie, tu as été toi aussi victime de cette injustice, terrassée par ce mal alors que tu cultivais le vivre sagement. Tu es partie alors que j'écrivais ces lignes, tu aurais tellement souhaité les lire, si curieuse et avide de connaissance que tu étais. J'espère avoir été digne de ton amitié, et te dédie ce manuscrit.

Ce travail n'aurait pu se réaliser sans toutes ces pierres et tous ces liants, et je vous en suis sincèrement redevable.

Table des Matières

Prédisposition héréditaire au cancer du sein et de l’ovaire : contexte général.....	11
Syndrome de prédisposition héréditaire au cancer du sein et de l’ovaire - Identification des premières bases moléculaires	11
Prise en charge des individus porteurs d’altération BRCA1 ou BRCA2	15
Critères d’adressage en consultation d’oncogénétique.....	18
Au-delà de BRCA1 et BRCA2 : élargissement du substratum moléculaire des prédispositions héréditaires au cancer du sein et de l’ovaire	21
Prédisposition héréditaire au cancer du sein et de l’ovaire - Approches par panel multi-gènes et contributions relatives	35
Variations privées - Phénotypes extrêmes.....	39
Objectifs de notre travail de Thèse.....	47
Partie I : Exomes soustractifs et comparatifs dans les cancers du sein et de l’ovaire de phénotype extrême.	48
Contexte et rationnel	48
Exomes soustractifs dans un trio parent / enfant affecté par un cancer de l’ovaire précoce.	48
Exomes soustractifs dans un trio parent / enfant affecté par un cancer du sein précoce.....	61
Exomes comparatifs dans une famille lourdement affectée par des cancers du sein ou du pancréas précoces.....	65
Discussion concernant les stratégies d’exomes soustractifs / comparatifs	67
Partie II : séquençage par panel de gènes impliqués dans la cancérogenèse de cohortes d’individus affectés par des tumeurs de phénotype extrême.....	69
Rationnel et Contexte.....	69
Patients et Méthodes	71
Résultats.....	78
Discussion	90
Partie III : recherche de variations délétères en mosaïque de TP53	94
Rationnel	94
Matériel et Méthode	95
Résultats.....	97
Discussion	105
Discussion Générale	109
Bibliographie.....	112
Annexes	126
Annexe : Protocole Capture 201 gènes	127

Liste des Figures

Figure 1 : Dommages et voies de réparation de l'ADN	12
Figure 2 : Risques cumulés de cancers du sein et de l'ovaire	14
Figure 3 : Paysage moléculaire des formes héréditaires de cancers du sein	39
Figure 4 : Situations conduisant à la survenue de variations <i>de novo</i>	42
Figure 5 : Identification des causes moléculaires de pathologies génétiques au cours du temps	43
Figure 6 : Interprétation d'un variant	46
Figure 7 : Stratégies déployées et inférences de la contribution de variants rares identifiés	47
Figure 8 : Protocole SureSelect Agilent et préparation des librairies d'ADN	50
Figure 9 : Formation des clusters - technologie Illumina	51
Figure 10 : Filtration des variants.....	54
Figure 11 : Séquençage du gène <i>INHBA</i> par méthode Sanger du cas index et des parents	54
Figure 12 : Données d'activité par méthode immunologique des protéines Activine et Inhibine en situation wild-type et mutée	59
Figure 13 : Filtration des variants.....	66
Figure 14 : BAM et séquences Sanger correspondant à la variation c.2459A>G d' <i>OGDH</i> ...	62
Figure 15 : BAM et séquence Sanger correspondant à la variation c.815T>C de <i>KLHL26</i> ...	63
Figure 16 : Arbre généalogique	65
Figure 17 : Modélisation du nombre de gènes inclus dans la capture en fonction des contraintes de profondeur et du nombre de <i>runs</i> nécessaires	75
Figure 18 : Visualisation sur ALAMUT et électrophorégramme SANGER de la variation <i>TP53</i> : c.[1024C=>T], p.[R342=/(R342X)].....	87
Figure 19 : Analyse par SNaPshot et analyse NGS du trio F281.....	101
Figure 20 : Analyse par SNaPshot et NGS du cas index F315	102
Figure 21 : Variations <i>TP53</i> en mosaïque sur électrophorégramme Sanger	102
Figure 22 : Confirmation par SNaPshot des mosaïques de <i>TP53</i>	103
Figure 23 : Perte d'hétérozygotie sur l'analyse de l'ADN tumoral	104

Liste des Tableaux

Tableau 1 : Synthèse des recommandations de prise en charge de la prédisposition héréditaire au cancer du sein et de l’ovaire	18
Tableau 2 : Risque de variation délétère <i>BRCA1/2</i> en fonction des critères cliniques	20
Tableau 3 : Critères de Chompret version 2015	23
Tableau 4 : Contribution des altérations <i>TP53</i> dans le cancer du sein précoce	26
Tableau 5 : Prévalence et <i>odds ratios</i> des principaux gènes candidats dans la prédisposition héréditaire au cancer du sein et de l’ovaire	34
Tableau 6 : Principales études analysant par panels de gènes des cohortes de cancers du sein	38
Tableau 7 : Variants rares non-synonymes identifiés dans la cohorte de cancers des ovaires précoces	57
Tableau 8 : Tests multiples évaluant l’hypothèse d’un enrichissement en variants rares chez les patients	58
Tableau 9 : Liste des 201 gènes par ordre alphabétique	74
Tableau 10 : Variants rares identifiés dans la cohorte de cancers du sein précoce	80
Tableau 11 : Associations entre variants rares et survenue d’un cancer précoce	89
Tableau 12 : Génotype et phénotypes associés des variations de novo de <i>TP53</i> identifiées	100
Tableau 13 : Variations en mosaïque de <i>TP53</i> identifiées.....	107

Prédisposition héréditaire au cancer du sein et de l'ovaire : contexte général

Syndrome de prédisposition héréditaire au cancer du sein et de l'ovaire - Identification des premières bases moléculaires

La description de familles remarquables par des agrégations de cancers du sein a conduit à suspecter et modéliser un syndrome de prédisposition héréditaire [1, 2], basé sur l'association de cas de cancers du sein chez des apparentés au premier degré (mère - fille, soeurs), la notion de cancers du sein bilatéraux, la survenue de cancer du sein chez des hommes et des âges d'apparition précoce [3]. L'identification de familles comportant de telles agrégations a permis dans un premier temps de suspecter un facteur génétique autosomique dominant [4], puis par de larges études de co-ségrégation (*linkage analysis*) d'identifier un microsatellite (*D17S579*) co-ségrégeant avec le risque de cancer du sein, situé au locus 17q21 [5] et proche du gène candidat *BRCA1* [6]. La présence de cas atteints de cancers de l'ovaire, initialement moins remarquables compte tenu de l'asymétrie des prévalences respectives entre le cancer du sein et de l'ovaire, a conduit à réduire l'hétérogénéité génétique de ces familles et ainsi renforcer la liaison au locus 17q21 de ces familles, définissant une **prédisposition héréditaire associant cancer du sein et de l'ovaire** (HBOC pour *Hereditary Breast and Ovarian Cancer*) [7, 8]. Les premières altérations de *BRCA1* ont été ainsi rapportées [9], de type variations non-sens ou perte de fonction, en accord avec le rôle de gène suppresseur de tumeur de *BRCA1*. Néanmoins, ces altérations de *BRCA1* n'expliquant que 45% de ces familles à haut risque de cancer du sein (80% en cas de cancers de l'ovaire associés), de nouvelles études de liaison ont été menées dans des familles agrégeant des cancers du sein mais *BRCA1* non muté, avec identification du locus 13q12-13 comportant le gène *BRCA2* [10].

BRCA1 et *BRCA2* sont deux gènes majeurs impliqués dans la réparation de l'ADN et plus particulièrement dans la réparation par recombinaison homologue [11]. *BRCA1* (MIM*113705), localisé en 17q21.31, comporte 22 exons codants et correspond à un grand gène de 110 kb encodant une protéine de 190 kD et 1863 acides aminés. Des épissages alternatifs sont décrits, notamment aboutissant à l'isoforme BRCA1-IRIS dont l'hyper-expression dans le cancer du sein triple négatif et le cancer de l'ovaire est associée à l'agressivité tumorale et à la résistance thérapeutique [12, 13]. *BRCA1* appartient à la famille des nucléo-phosphoprotéines, et interagit avec les effecteurs de la réparation de l'ADN (complexe BASC pour *BRCA1-associated genome surveillance complex*) en réponse à l'activation des protéines *sensor* des dommages à l'ADN (ATM, ATR). Par son extrémité N-terminale, *BRCA1* exerce également un rôle dans la régulation transcriptionnelle par interaction avec l'ARN polymérase II, et par son extrémité C-terminale un rôle dans le remodelage chromatinien par une interaction avec les histones desacetylases. Le rôle de *BRCA1* dans la voie de la réparation de l'ADN est complexe. *BRCA1* est phosphorylé par ATM, ATR et CHK2 en présence de cassures double-brin de l'ADN. Ainsi activé, *BRCA1* est relocalisé vers les régions d'ADN altérées et impliquées dans les deux mécanismes distincts permettant la

réparation des cassures double-brins de l'ADN : NHEJ (*non-homologous end joining*, système de réparation non fidèle) et HR (*homology-directed recombination*) (Figure 1). Le mécanisme de recombinaison homologue, garantissant une réparation fidèle par l'identification de séquences à fortes homologies, ne peut-être mis en jeu qu'en phase S/G2 afin de permettre l'utilisation de la chromatide soeur comme matrice. BRCA1 interagit tant avec le complexe MRN (MRE11-RAD50-NBN) impliqué dans le mécanisme NHEJ qu'avec RAD51 et BRCA2 impliqués dans la recombinaison homologue. BRCA1 jouerait ainsi un rôle d'orchestrateur et de régulateur entre ces deux voies, notamment par la régulation de 53BP1 [14]. L'activité ubiquitine-ligase de BRCA1 régule également FANCD2 et participe ainsi également à la régulation de la réparation des liaisons covalentes inter-brins (ICL pour *interstrand DNA crosslinks*).

BRCA2 (MIM*600185) appartient au complexe de Fanconi (*FANCD1*) et est localisé en 13q13.1. BRCA2 comporte 27 exons, avec un grand exon central (exon 11) comme BRCA1, représentant plus de 85kb et codant pour une protéine de 3418 acides aminés de 384kDa. BRCA2 est un acteur majeur de la recombinaison homologue, par son interaction avec RAD51 permettant la liaison de ce dernier sur les fragments d'ADN simple brin afin d'activer la réparation. BRCA2 est recruté et activé par interaction indirecte avec BRCA1 via PALB2 [15].

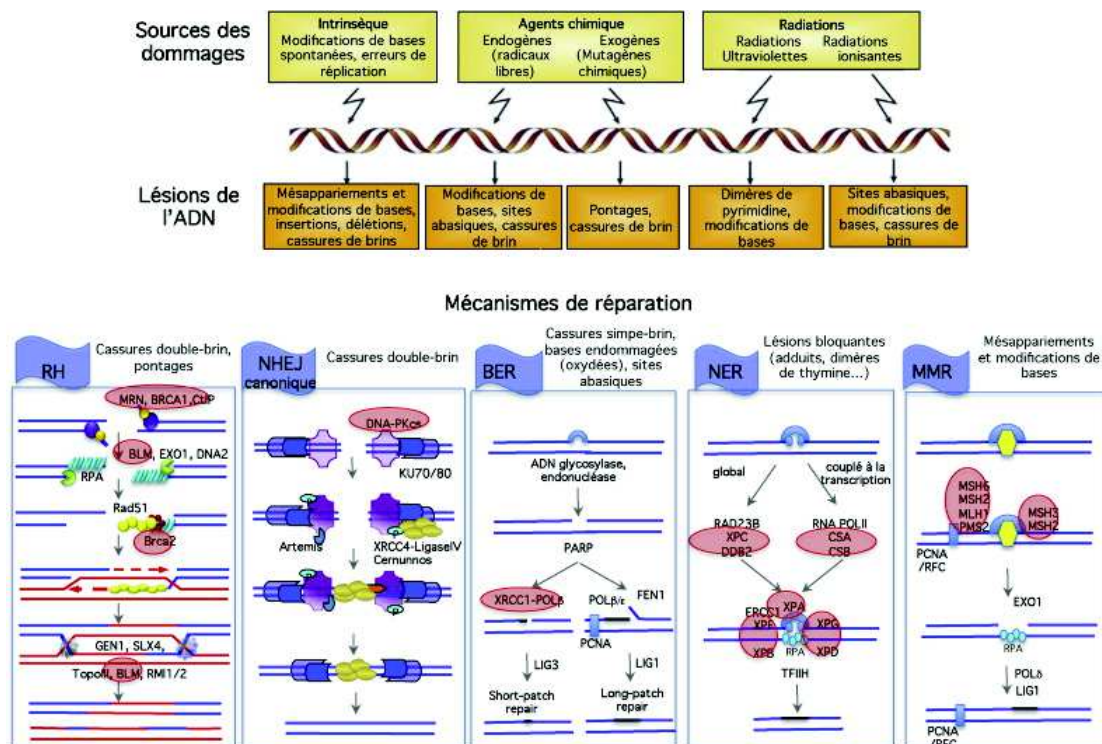


Figure 1 : Dommages et voies de réparation de l'ADN. D'après Rass *et al*, Cancer/Radiothérapie 2012 [11]

Les altérations de *BRCA1* et *BRCA2* sont répertoriées dans des bases de données internationales : BIC, LOVD, UMD, ClinVar, ENIGMA [16-20]. Le projet *BRCA Exchange* [21] agrège l'ensemble des données issues du BIC, LOVD et ClinVar afin de proposer un portail unique d'interrogation des variants de *BRCA1* et *BRCA2*, annotés et classifiés selon la classification de l'ACMG/AMP (*American College of Medical Genetics and Genomics / Association of Molecular Pathology*) [22]. En cohérence avec le statut de gène suppresseur de tumeur, les altérations délétères de *BRCA1* et *BRCA2* sont principalement des altérations pertes de fonction, de type mutation stop, *Indels* rompant le cadre de lecture (*frameshift*), variations d'épissage, ou réarrangements de grande taille. Les altérations correspondent principalement à des variations nucléotidiques ponctuelles, tel qu'illustré dans la base UMD-*BRCA1* avec, pour 2495 variants différents rapportés, les proportions suivantes : réarrangements de grande taille (≥ 1 exon) $n=77$ (3%), autre CNV (< 1 exon) $n=504$ (20%), variations introniques à proximité des bornes exoniques (± 10 paires de bases) $n=133$ (5%), indels $n=39$ (1,5%), variations introniques profondes $n=606$ (25%) et enfin variations exoniques ponctuelles $n=1136$ (45%), principalement de type mutations stop.

La principale difficulté d'interprétation réside dans les variations de type faux-sens ou synonyme, dont l'interprétation selon les critères de l'ACMG/AMP ne permet pas de classer en variation bénigne ou probablement bénigne (classe 1 - 2) ou variation délétère ou probablement délétère (classe 4 - 5), définissant ainsi la catégorie des **variants de signification inconnue** ou VSI (classe 3). Ainsi, la base de données UMD-*BRCA1* rapporte sur les 2495 variants différents, 1356 VSI pour 891 variants pathogènes de classe 5 [18]. Les proportions sont similaires pour *BRCA2* avec, pour 3454 variants différents rapportés, 2181 VSI pour 968 variants pathogènes de classe 5. *BRCA2* correspond ainsi au gène humain le plus concerné par ces VSI [23]. Des études de co-ségrégation (COVAR, NCT01689584) recherchent des arguments épidémiologiques en faveur d'une pathogénicité ou bénignité. L'interprétation d'un éventuel effet sur l'épissage des VSI proches des bornes exoniques repose sur des études fonctionnelles, par analyse des transcrits sur prélèvement PAXgene des patients porteurs et test *ex vivo* de type minigène [24]. Les prédictions bio informatiques permettent de suspecter d'éventuels effets sur l'épissage, et ainsi de prioriser les analyses fonctionnelles de ces variants [25, 26]. En raison de la régulation complexe de l'épissage mettant en jeu notamment des séquences exoniques (ESE/ESS pour *Exonic Splicing Enhancer* ou *Silencer*), les variations exoniques faux-sens ou synonymes doivent également s'interpréter en fonction d'un potentiel effet sur l'épissage [27]. Les larges exons centraux de *BRCA1* et de *BRCA2* (exon 11) sont l'objets d'épissages alternatifs physiologiques et d'une régulation complexe, rendant délicate l'interprétation de l'éventuel impact sur l'épissage des variants présents dans ces exons [28]. D'importants travaux ont été menés afin de rapporter et classer ces variants de signification inconnue, avec une réduction des discordances inter-laboratoires dans l'interprétation de ces variants [29]. Les enjeux de cette classification sont en effet majeurs puisque conditionnant la prise en charge des familles, tel qu'illustré par une étude rétrospective ré-interprétant les variants de classe 3 rapportés par un acteur privé américain (*Myriad Genetics*), avec, pour 7,7% des variants de classe 3, une re-classification en classe 1-2

dans 91% des cas mais surtout une re-classification en variants de classe 4-5 dans 9% des cas, impliquant un défaut de prise en charge pour ces familles [30].

Les risques associés à ces altérations et les pénétrances respectives ont été récemment stabilisés (Figure 2) [31]. Ainsi, concernant les porteurs d'altérations *BRCA1*, le risque mammaire cumulé à 80 ans est de 72%, avec une augmentation rapide du taux d'incidence annuel à partir de 30 ans (23,5/1000 personnes-années) et le déclin de ce taux uniquement après 70 ans. Pour les altérations de *BRCA2*, le risque mammaire cumulé à 80 ans est similaire (69%) mais avec un décalage de 10 ans de l'âge d'augmentation de l'incidence (10,8/1000 personnes-années entre 31 et 40 ans puis 27,5/1000 personnes-années entre 41 et 50 ans). Concernant le risque ovarien, il est davantage lié à *BRCA1* avec un risque cumulé à 80 ans de 44% versus 17% pour *BRCA2*. L'incidence augmente à partir de 41 ans pour *BRCA1*, et de même s'observe un décalage de 10 ans pour *BRCA2*. Enfin concernant le risque de cancer du sein controlatéral, l'incidence est majeure d'autant plus que l'âge de survenue du premier cancer est précoce, avec des taux annuels incidents entre 20 et 30/1000 personnes-années pour *BRCA1*, 15 à 20/1000 personnes-années pour *BRCA2*.

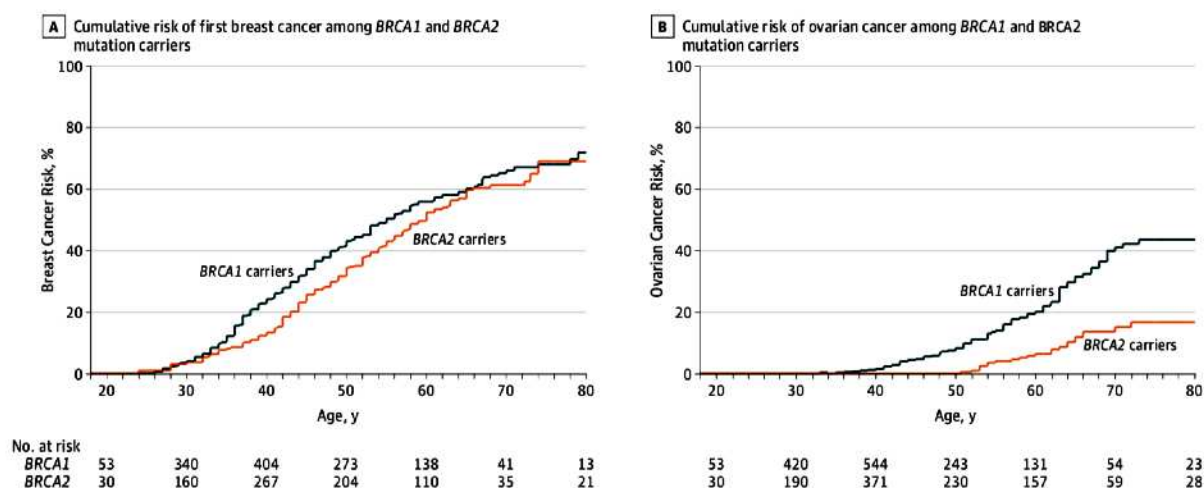


Figure 2 : Risques cumulés de cancers du sein et de l'ovaire (Kuchenbaecker *et al.* 2017) [31]

Le risque de cancer du sein chez les hommes porteurs d'altérations constitutionnelles est davantage lié à *BRCA2* qu'à *BRCA1*, avec un risque cumulé respectivement de 6,8% et 1,2%, comparativement à 0,1% dans la population générale [32]. En revanche, les hommes sont concernés par un sur-risque de cancer prostatique, avec des *odds ratios* estimés à 2,64 (IC95% 2,03-3,47) pour *BRCA2* et 1,35 (IC95% 1,03-1,76) pour *BRCA1* [33]. Le spectre tumoral de la prédisposition héréditaire au cancer du sein et de l'ovaire est enfin enrichi d'un sur-risque de cancer du pancréas, avec des risques relatifs standardisés de 2,55 et 2,13 respectivement pour *BRCA1* et *BRCA2* [34]. D'autres sur-risques ont été évoqués (colon, mélanome, voies biliaires) [35, 36].

Sur un plan histopathologique, il n'existe pas de forme pathognomonique de cancers du sein ou de l'ovaire indicative d'une origine génétique. De ce fait, la suspicion d'une forme héréditaire sera basée sur l'anamnèse familiale et individuelle, à la recherche d'arguments d'agrégation familiale, de précocité et de multiplicité des cancers. Néanmoins, un enrichissement de formes héréditaires est noté dans les formes triples négatives de cancers du sein [37, 38], de haut grade de même que dans les carcinomes médullaires mammaires [39]. Dans une large étude de plusieurs cohortes de cancers du sein dont l'histopathologie était disponible, le grade 3 apparaissait davantage prédictif de la présence d'une variation délétère que le statut des récepteurs hormonaux, avec néanmoins l'association grade 3 triple négative hautement prédictive d'une altération de *BRCA1*, quel que soit l'âge (inférieur ou supérieur à 50 ans) [40]. Le caractère triple négatif apparaît également corrélée à un risque relatif de 5,65 de variation délétère de *BRCA1* comparativement à la population de cancer du sein non triple négative [41]. Les tumeurs mammaires *BRCA2* ont néanmoins un profil commun aux tumeurs sporadiques, de grade intermédiaire ou de haut grade, et de profil luminal. Les observations issues des consultations en oncogénétique tendent néanmoins à retenir des profils récurrents de cancers du sein en situation d'altérations *BRCA1* ou *BRCA2* identifiées avec : développement rapide des tumeurs détectées par autopalpation, grade SBR II ou III, indice de prolifération (Ki67) élevé. Les carcinomes ovariens sont majoritairement de type séreux (67%), les formes endométrioïdes (12%), mucineuses et à cellules claires étant néanmoins possibles [42].

Prise en charge des individus porteurs d'altération BRCA1 ou BRCA2

La prise en charge des individus porteurs d'altération *BRCA1* ou *BRCA2* prend de ce fait en compte les risques mammaires et ovariens en proposant, de manière individualisée à l'histoire personnelle et familiale, des stratégies de réduction du risque. Concernant le risque de cancer du sein, les recommandations publiées par l'Institut National du Cancer (INCa) synthétisent les données de la littérature (Tableau 1) [43]. L'examen clinique par palpation mammaire et des creux axillaires est pour la détection du cancer du sein une modalité de dépistage peu sensible, mais conservant toute sa place avant la mise en place de la surveillance par imagerie (25-30 ans) puis en alternance avec le suivi radiologique afin d'instaurer une surveillance semestrielle. L'examen central de la surveillance radiologique est l'IRM mammaire, compte tenu de sa sur-performance par rapport à la mammographie et à l'échographie, en particulier chez les femmes de moins de 50 ans à forte densité mammaire (BI-RADS 3 et 4). Selon les études, le gène altéré et l'âge de la femme, la sensibilité de l'IRM varie de 66% à 92% pour des spécificités respectives de 85 à 95% [44-48]. La triple modalité de dépistage associant sur un rythme annuel la réalisation de l'IRM mammaire, mammographie et échographie, offre la sensibilité de détection la plus élevée, notamment pour la détection des formes *in situ* échappant éventuellement à la détection par IRM [48]. L'âge d'initiation est classiquement 30 ans, sauf cas plus précoce dans l'histoire familiale avec anticipation de 5 ans dans ce cas de l'âge d'initiation par rapport au cas familial le plus précoce. Au-delà de 65 ans, la place de l'IRM est plus débattue, compte tenu de la réduction de

la densité mammaire permettant une détection optimale par mammographie. Compte tenu du risque de cancers du sein radio-induits dans cette population précocement exposée à la mammographie et présentant une altération constitutionnelle de la réparation de l'ADN à même de la rendre plus sensible aux effets délétères des radiations ionisantes, le compromis à ce jour est l'initiation de la mammographie à partir de 30 ans et la limitation à un seul cliché oblique externe en technique numérique plein champ tant que l'IRM est associée. La démonstration de l'efficacité de cette surveillance renforcée est largement focalisée sur le taux de détection des cancers dans cette population à risque, avec cependant moins de données concernant la réelle réduction du sur-risque de mortalité. L'approche par surveillance renforcée semble moins effective dans la population avec altération *BRCA1* [49] comparativement à la population *BRCA2* [50]. Dans cette dernière étude, il est démontré une survie spécifique à 10 ans sans cancer du sein de 100% pour la population bénéficiant d'une surveillance mammaire par IRM comparativement à 85,5% pour la population surveillée par mammographie seule et 74,6% pour la population non surveillée. La réduction du risque de décès est démontrée comme résultant d'une plus grande précocité du diagnostic, avec notamment moins de formes évoluées (>2cm et/ou avec atteinte ganglionnaire).

Le moindre bénéfice de la surveillance renforcée observée dans la population *BRCA1*, le risque de cancer de l'intervalle lié à la plus grande incidence de cancer du sein de haut grade à développement subaigu, la préférence exprimée par les individus porteurs amènent à considérer l'option chirurgicale par mastectomie bilatérale prophylactique chez les individus sains à risque, ou contralatérale prophylactique chez les individus déjà affectés. Le risque de cancer de l'intervalle est limité par l'usage de l'IRM, et observé quasi exclusivement chez les femmes *BRCA1* mutées [51, 52]. La mastectomie prophylactique apparaît intuitivement comme l'approche la plus sécurisante, avec des risques de cancers du sein résiduels inférieurs à 10% à 70 ans [53-55]. Cette option chirurgicale, qui est aujourd'hui une recommandation pour les formes porteuses d'une variation délétère de *BRCA1*, a pénétré plus tardivement en France que dans les pays Anglo-Saxons, en raison de facteurs socio-culturels. Le risque de cancer du sein résiduel doit néanmoins être explicité aux patientes dans une démarche d'information éclairée, et est corollaire aux techniques de préservation de la plaque aréolo-mamelonnaire et/ou de l'étui cutané permettant un meilleur résultat esthétique et fonctionnel après mastectomie - reconstruction. Pour des raisons éthiques, aucune donnée prospective randomisée ne permet d'établir formellement un gain de survie en faveur de l'approche chirurgicale, les études rapportées comparant le devenir d'individus à haut risque faisant le choix ou non d'une approche chirurgicale, ce choix étant inévitablement biaisé par de multiples facteurs dont l'agressivité de la variation délétère (pénétrance, phénotypes). Les études observationnelles [56, 57] proposent néanmoins une réduction du risque de l'ordre de 10 à 20%, davantage corrélée à la population *BRCA1* que *BRCA2*. Les modèles simulés permettent néanmoins d'extrapoler un bénéfice de survie sans cancer du sein, de l'ordre de 8,7% de gain de survie à 80 ans pour une mastectomie préventive effectuée à 25 ans [58], et avec une réduction graduelle de ce bénéfice lorsque la mastectomie est réalisée plus tardivement [58-60]. Pour les sujets atteints de cancer du sein, la

place de la mastectomie thérapeutique plutôt que d'un traitement conservateur est moins ambiguë, même si le risque accru de récurrence ou plutôt de second cancer homolatéral observé en cas de traitement conservateur ne s'accompagne pas d'une détérioration de la survie globale ou spécifique [61, 62]. La réduction du risque de mortalité après mastectomie contralatérale prophylactique pour les patientes porteuses de variations délétères ayant déjà présenté un cancer du sein unilatéral est en revanche démontrée, avec une réduction du risque de mortalité de 51% [63, 64], mais principalement pour les patientes de moins de 40 ans présentant un cancer du sein de faible grade, non triple-négatif et n'ayant pas relevé d'une chimiothérapie adjuvante [64]. Cette dernière observation implique que pour les patientes ayant présenté un premier cancer agressif, le risque vital est dominé par le risque de récurrence métastatique de ce premier cancer, et conduit à envisager une période minimale d'observation avant proposition d'une mastectomie contralatérale prophylactique.

L'annexectomie bilatérale prophylactique à partir de 40 ans est la mesure de réduction du risque la plus consensuelle permettant efficacement, en l'absence d'alternative, la prévention du cancer des ovaires ou des trompes. La réduction de la mortalité globale est ainsi de 68%, légèrement supérieure en cas d'historique de cancer du sein et significative tant pour les patientes mutées *BRCA1* que *BRCA2* [65]. La réduction du risque de cancers des annexes est majeure (80%) avec un risque résiduel cumulé à 70 ans de 9% pour les patientes avec altération de *BRCA1* et 2% pour celles impliquant *BRCA2* [56, 59]. L'âge de réalisation tient compte des éventuels projets parentaux, de l'histoire familiale et personnelle, et du gène altéré puisque l'âge incident pour le cancer de l'ovaire est plus tardif de 10 ans en situation de variation délétère de *BRCA2*. L'intérêt de cette mesure prophylactique est double, puisqu'il est établi que l'annexectomie prophylactique réduit significativement le risque mammaire chez la femme non ménopausée. Dans la méta-analyse de Rebbeck *et al*, la réduction du risque mammaire est de 51%, égale entre les femmes mutées *BRCA1* et *BRCA2* [66]. Néanmoins, si ce bénéfice est établi, l'amplitude du bénéfice, les populations bénéficiaires (ménopausées ou non, gène altéré) sont contradictoires dans la littérature [56, 59, 67].

Les approches hormonales préventives sont plus débattues, les études d'hormono-prévention par les *SERM* ou inhibiteurs de l'aromatase étant positives uniquement du point de vue de la réduction de l'incidence du cancer du sein dans les populations de femmes à risques moyens ou élevés, sans traduction sur la mortalité globale ni spécifique [68, 69]. Aucune donnée n'est disponible en prévention primaire pour les patientes avec altération de *BRCA1* ou *BRCA2*. L'efficacité de l'hormonothérapie en prévention secondaire chez les patientes mutées est en revanche similaire à la population des cancers du sein sporadiques.

	Âge début	BRCA1	BRCA2
Risque Mammaire	25 ans	Examen clinique semestriel et auto-palpation	
	30 ans	Mastectomie bilatérale prophylactique +++	Surveillance trimodale par IRM - Mammographie - Echographie
		Option : Surveillance trimodale par IRM - Mammographie - Echographie	Option : Mastectomie bilatérale prophylactique
Risque Ovarien	40 - 45 ans	Annexectomie bilatérale prophylactique	
Risque Prostatique	40 ans	Examen clinique - dosage annuel PSA	

Tableau 1 : Synthèse des recommandations de prise en charge de la prédisposition héréditaire au cancer du sein et de l'ovaire

Critères d'adressage en consultation d'oncogénétique

Bien que par leur prévalence, les syndromes de prédisposition aux cancers représentent la première pathologie génétique, bien avant les pathologies du développement, la majorité des cancers demeurent d'origine somatique. Si environ 20% des patientes atteintes de cancer du sein présentent une histoire familiale compatible avec une forme héréditaire, la forte prévalence de cette pathologie dans la population générale implique statistiquement une probabilité non nulle que plusieurs femmes au sein d'une même famille développent des cancers du sein, indépendamment d'un facteur héréditaire. De ce fait, l'estimation de la prévalence des formes héréditaires de cancers du sein est plutôt estimée proche de 10%. Par ailleurs, la reconnaissance formelle d'une origine héréditaire suppose la mise en évidence de l'altération constitutionnelle causale, impliquant de nombreux biais dans l'estimation de cette prévalence tels que des biais de recrutement (population étudiée, enrichie ou non en formes précoces de cancers du sein ou en patientes présentant une histoire familiale compatible, exhaustivité de l'adressage en consultation d'oncogénétique), les biais d'analyse et notamment l'absence de connaissance exhaustive de

l'ensemble des déterminants moléculaires à l'origine de ces formes héréditaires. Néanmoins, en restreignant cette étude de la prévalence aux deux gènes majeurs impliqués dans la prédisposition héréditaire au cancer du sein et de l'ovaire à savoir *BRCA1* et *BRCA2*, et dans une population anglo-saxonne de 1435 patientes atteintes de cancers du sein avant 55 ans, 2% de formes héréditaires avaient été documentées (0,7% et 1,3% de variations délétères de *BRCA1* et *BRCA2* respectivement) [70]. On retrouve dans cette étude ancienne l'enrichissement en prévalence de variations constitutionnelles des formes les plus précoces et/ou associées à une histoire familiale significative (12% de variations dans les formes de moins de 35 ans, 4% en cas d'antécédent familial au premier degré de cancer du sein ou de l'ovaire). Certaines entités géographiques ou résultant d'un brassage génétique restreint présentent une sur-prévalence de variations affectant *BRCA1* ou *BRCA2*, notamment la population ashkénaze [71], mais également polonaise [72].

Différents algorithmes prenant en compte les âges de survenue, les formes histologiques et les histoires familiales, définissent des probabilités d'identification d'une variation délétère, principalement de *BRCA*, ainsi que des probabilités de survenue d'un cancer du sein ou de l'ovaire. Les algorithmes BRCAPRO [73] ou BOADICEA [74] supposent la détermination, assistée informatiquement, de l'histoire familiale par l'établissement de l'arbre généalogique. Le score de Manchester [75], basé sur une pondération affectée à l'existence au sein d'une famille de cancers du sein chez la femme, chez l'homme, de cancers de l'ovaire, de cancers de la prostate ou de cancers pancréatiques, modulée par les âges d'apparition, présente l'avantage de sur-performer le score BRCAPRO et de pouvoir être calculé sans connaissance exacte des liens de parenté. Ce score est calibré pour que chaque incrément de 10 points soit associé à une augmentation de probabilité de détecter une variation délétère de *BRCA1* ou *BRCA2*. Le score d'Eseinger [76] reprend la méthodologie du score de Manchester sur un mode simplifié, mais consiste davantage à légitimer l'intérêt d'une consultation en oncogénétique plutôt qu'à déterminer une probabilité d'altération génétique, en retenant la probabilité minimale de 10% de détection d'une variation délétère de *BRCA* conformément aux recommandations de l'ASCO [77]. En effet, les probabilités d'identification d'une altération *BRCA*, lorsque déterminées à partir de critères individuels ou familiaux mutuellement exclusifs, sont faibles (Tableau 2) [78]. Ainsi, la précocité avant 36 ans d'un cancer du sein, lorsque seul critère, aboutit à un taux de détection de variants délétères inférieur à 10% (7,9%, IC95% [3,6 - 12,2]). De même, la survenue d'un cancer du sein non triple négatif et ce quelque soit l'âge de survenue, chez un seul individu mais bien plus encore, une agrégation de cancers du sein chez plusieurs apparentés sans critère de précocité, ne s'associent pas à des *odds ratio* significatifs. Le caractère bilatéral même avant 50 ans d'un cancer du sein n'apparaît pas, de même, significatif. On retrouve finalement dans cette étude les critères à fort poids statistique que sont la présence d'un cancer des ovaires précoce (<46 ans) ou associé à un cancer du sein chez le même individu ou chez un apparenté, et le caractère triple négatif du cancer du sein lorsque l'âge de survenue est relativement précoce (43 ans).

Performance of mutually exclusive selection criteria.

Selection criteria (probands/families with only one criterion fulfilled)	Mutation carriers				OR	95% CI	P	
	N	DR % (95% CI)	BRCA1	BRCA2				
One case affected								
1) BrCa <36y	151	12	7.9 (3.6–12.2)	3	9	0.42	0.2–0.82	<0.001
2) BrCa ≥36y + OvCa ≥46y	22	7	31.8 (16.3–52.7)	6	1	3.05	1.00–8.32	0.015
3) BrCa ♂ any age	26	1	3.8 (0.68–18.9)	0	1	0.24	0.005–1.49	0.12
4) OvCa <46y	15	4	26.6 (10.9–52.0)	1	3	2.30	0.52–8.07	0.15
5) BBC <50y (first BrCa ≥36y)	7	2	28.5 (8.2–64.1)	2	0	2.50	0.23–15.7	0.26
6) TNBC <43y (BrCa ≥36y)	2	1	50.0 (9.4–90.6)	1	0	6.25	0.08–492.0	0.14
Two 1st-degree relatives								
7) BrCa ≥36y and <50y	64	10	15.6 (8.7–26.4)	4	6	1.14	0.49–2.43	0.72
8) BrCa ≥36y and <50y + BBC ≥50y	21	1	4.8 (0.8–22.6)	1	0	0.23	0.01–1.93	0.21
9) BrCa ≥36y and <50y + OvCa ≥46y	22	7	31.8 (16.4–52.7)	6	1	3.05	1.00–8.32	0.01
10) BBC ≥50y + OvCa ≥46y	5	4	80.0 (37.6–96.4)	2	2	26.0	2.49–1283.1	<0.001
11) OvCa ≥46y	11	4	36.4 (15.2–64.6)	3	1	3.66	0.76–14.82	0.03
12) BrCa ≥36y and <50y + BrCa ♂ (female proband)	3	0	0.0	0	0	N.A.	N.A.	N.A.
13) BBC ≥50y + BrCa ♂ (female proband)	1	0	0.0	0	0	N.A.	N.A.	N.A.
Three 1st-degree relatives								
14) ≥3 BrCa any age (n.1 and n.7 excluded)	125	9	7.2 (3.8–13.1)	4	5	0.39	0.16–0.82	<0.001
15) 2 BrCa ≥50y + OvCa ≥46y	9	4	44.4 (18.9–73.3)	4	0	5.15	0.99–24.5	0.01
16) 2 OvCa ≥46y + BrCa ≥50y	5	2	40.0 (11.8–76.9)	2	0	4.20	0.34–37.2	0.09
Total	489	68	13.9	39	29			

The 489 probands and families reported in the table fulfill only one of the selection criteria. Age thresholds in years (y) are indicated for each criterion; the mutation detection rate (DR) of each group and the significant odds ratios (OR) are highlighted in bold. BrCa, breast cancer; OvCa, ovarian cancer; BBC, bilateral breast cancer; TNBC, triple negative breast cancer; VUS, variants of uncertain significance; CI, confidence interval; N.A., not applicable.

Tableau 2 : Risque de variation délétère de BRCA1/2 en fonction des critères cliniques (Azzollini *et al*, Eur. J. Int. Medicine 2016) [78]

Ainsi, afin de garantir un taux de détection minimal de 10% de variants pathogènes de *BRCA*, la plupart des Laboratoires, à l'instar du Laboratoire de Biologie et de Génétique du Cancer du Centre François Baclesse de Caen réalisant les analyses moléculaires du Grand Ouest, retient les critères suivants :

• Critères individuels :

- cancer du sein avant 36 ans
- cancer médullaire du sein avant 61 ans
- cancer du sein triple négatif avant 51 ans
- cancer du sein de l'homme avant 71 ans
- cancer de l'ovaire avant 61 ans

• Critères familiaux :

- 2 cas de cancers du sein (dont l'un avant 51 ans et l'autre avant 71 ans) apparentés aux 1^{er} ou 2nd degré
- 3 cas de cancers apparentés aux 1^{er} ou 2nd degré dont l'un avant 61 ans
- cancer du sein avant 51 ans associé à un apparenté atteint de cancer prostatique avant 61 ans, de cancer du pancréas avant 61 ans ou de cancer des ovaires avant 71 ans.

Néanmoins, force est de constater, et probablement regretter, qu'en 2019 les indications de réalisation d'un test génétique ont fortement dérivé de par l'avènement du séquençage de nouvelle génération et la simplification de l'analyse génétique. Ainsi, certains centres au niveau international promeuvent la proposition d'analyse génétique à toute femme atteinte de cancer du sein ou de l'ovaire. Cela conduit à l'identification d'un nombre croissant de variations de

signification inconnue et pose des questions de pertinence tant clinique que médico-économique. Bien plus, cette dérive menace la qualité de l'interprétation des variants, en polluant les bases de données de variants identifiés dans des phénotypes peu évocateurs de prédisposition héréditaire. A l'opposé, afin de garantir la pertinence de l'indication et de réguler ainsi la demande de consultation à même de prioriser et de garantir l'accès dans les meilleurs délais aux demandes urgentes (avec impact thérapeutique dans le contexte d'un cancer du sein en cours de prise en charge - cancer de l'ovaire avec indication d'inhibiteur de PARP), l'Unité de Génétique Clinique de Rouen a mis en place un filtre préalable à toute demande par le biais des entretiens téléphoniques de préparation (ETP). Il s'agit de pré-consultations assurées par des conseillères en génétique, permettant d'élaborer par téléphone avec le patient l'arbre généalogique, de préciser le contexte et de récupérer en amont de la consultation les autorisations d'accès aux dossiers médicaux du patient et de ses apparentés. Cette expérience a ainsi largement rempli son objectif de régulation et a démontré son bénéfice médico-économique [79]. Dans l'actualisation récente de notre activité proposée aux Assises de Génétique Humaine et Médicale 2020, nous décrivons ainsi 36% de consultations inappropriées (pas d'indication pertinente, mauvais cas index) et 10% de consultations jugées prioritaires (rendez-vous souhaitables et organisés dans les deux semaines).

Au-delà de BRCA1 et BRCA2 : élargissement du substratum moléculaire des prédispositions héréditaires au cancer du sein et de l'ovaire

Comme indiqué précédemment, les altérations de *BRCA1* et *BRCA2* n'expliquent qu'une fraction (15%) des formes familiales même les plus évocatrices [80], conduisant à supputer l'existence d'autres altérations génétiques causales. Néanmoins, cette quête, bâtie sur le paradigme qu'une agrégation familiale est probablement résultante d'une altération génétique et que réciproquement, les formes sporadiques sont peu évocatrices, est probablement biaisée. Ainsi, une étude de simulations épidémiologiques suggère qu'une proportion importante de formes familiales (30%) résulte d'un facteur chance plutôt que d'un facteur héréditaire [81]. A l'inverse, dans une population ashkénaze, la majorité des apparentés de sujets porteurs d'altérations de *BRCA1* ou *BRCA2* sont indemnes, avec de ce fait une présentation sporadique apparente [82]. Ces limites expliquent en partie l'échec des études de liaison renouvelée afin d'identifier un éventuel troisième locus (*BRCA3*) impliqué dans les formes familiales *BRCA1/2* négatives. Ainsi furent proposés, sans grande conviction, les loci 4q [83], 13q21 [84, 85], le locus du gène *ESR1* [86], 2q32 [87], 8p12-p22 [88-90].

En parallèle de ces études fondées sur une sélection de cas familiaux évocateurs, d'autres approches se sont fondées sur l'existence d'un dénominateur commun histopathologique ou moléculaire chez ces familles *BRCAx*. Ainsi, les comparaisons cytogénétiques des gains et pertes observés au niveau des tumeurs mammaires *BRCAx* et sporadiques, suggèrent un enrichissement d'évènements communs aux tumeurs *BRCA2* et un profil distinct des tumeurs

sporadiques [91-93]. A l'inverse, la recherche de profils immunohistochimiques au niveau des tumeurs BRCAx a échoué à identifier une *clusterisation* unique, suggérant davantage une hétérogénéité et un modèle polygénique [94]. Néanmoins, par une approche transcriptomique et une classification supervisée, un sous-groupe de tumeurs BRCAx, plutôt de sous-type luminal A, apparaît distinct des formes sporadiques et impliquer la voie de la réparation de l'ADN [95].

Cette voie de la réparation de l'ADN est en effet largement impliquée dans la prédisposition héréditaire au cancer du sein et de l'ovaire, avec plusieurs gènes contribuant à cette voie biologique et impliqués dans les présentations familiales *BRCA1/2* négatives.

TP53

Description initiale du syndrome de Li-Fraumeni

Le syndrome de Li-Fraumeni (LFS) a été identifié sur des bases épidémiologiques, par l'observation dès 1969 d'une prédisposition héréditaire au cancer, de transmission autosomique dominante, avec agrégations de sarcomes, cancers du sein et d'autres néoplasies, marquées par une grande précocité [96]. La définition stricte de ce syndrome est stabilisée en 1988 : **cas index atteint de sarcome avant l'âge de 45 ans associé à au moins deux apparentés (de premier ou second degré mais au minimum un apparenté au premier degré) atteints de cancer quel qu'en soit le type avant 45 ans ou de sarcome quel qu'en soit l'âge** [97]. Les sarcomes affectent davantage les tissus mous, avec une sur-représentation des rhabdomyosarcomes de type anaplasiques. Des tumeurs cérébrales (gliomes mais aussi des tumeurs rares telles que tumeurs des plexus choroïdes, médulloblastomes, épendymomes), des hémopathies malignes et des corticosurrénalomes définissent le panel des tumeurs présentées par les enfants / adolescents. Les cancers du sein de la femme très jeune apparaissent dès ce registre *princeps* comme une composante de ce syndrome. Les cancers primitifs multiples et la susceptibilité aux radiations ionisantes des tissus exposés sont également illustrés dans cette observation initiale. Outre les classifications de Birch [98] et de Eeles [99] qui définissent des formes *LFS-like* en adaptant les critères de la définition initiale, les critères de Chompret proposent une relecture complète du syndrome de Li-Fraumeni basée sur l'observation des phénotypes des patients et familles porteurs d'une altération de *TP53*. En effet, si les pathologies sus-mentionnées constituent les tumeurs emblématiques du syndrome de Li-Fraumeni, de nombreuses observations de familles conduisent à souligner et à intégrer dans cette définition syndromique des situations marquées par d'autres cancers pédiatriques ou des tumeurs de précocité inhabituelle chez l'adulte, au premier chef desquelles figurent le cancer du sein, **et ce en l'absence d'histoire familiale évocatrice**. Cette évolution conceptuelle a conduit à proposer le terme de cancers associés à une variation délétère constitutionnelle de *TP53* (*hTP53-rc* pour *hereditary TP53-related cancers*). Les **critères de Chompret** dont l'actualisation a été pilotée par l'équipe de Rouen ces dernières années [100, 101] proposent ainsi les critères suivants aux côtés de la définition classique : (i) tumeurs primitives multiples dont deux appartenant au spectre tumoral emblématique du syndrome LFS et survenant pour l'une avant 46 ans, quelque soit

l'histoire familiale, (ii) tumeurs rares évocatrices parmi lesquelles figurent le corticosurréalome, les tumeurs des plexus choroïdes, les rhabdomyosarcomes embryonnaires de type anaplasique, quelque soit l'histoire familiale, (iii) les cancers du sein de la femme jeune avant 31 ans quelque soit l'histoire familiale (Tableau 3). La survenue de cancers dans un champs d'irradiation antérieur, notamment avec un délai de latence court, constitue également une porte d'entrée pertinente dans la reconnaissance d'un syndrome de Li-Fraumeni, incluse dans le contexte des tumeurs primitives multiples.

Critères nosologiques		Taux de détection <i>TP53</i> (%)
Cas index avec tumeur du spectre LFS avant 46 ans et apparenté au 1 ^{er} /2 ^{ème} degré atteint avant 56 ans		29
Tumeurs primitives multiples incluant 2 cancers du spectre LFS et l'une avant 46 ans		ND
Tumeurs rares	Corticosurréalome de l'enfant	50
	Corticosurréalome de l'adulte	18
	Tumeurs des plexus choroïdes	50
	RMS anaplasique	73
Cancer du sein avant 31 ans		6

Tableau 3 : Critères de Chompret version 2015 [101] : ND : Non déterminé

Les bases moléculaires du syndrome de Li-Fraumeni sont posées en 1990 par l'identification, chez des sujets atteints, de variations de *TP53* dans les domaines de liaison à l'ADN hautement conservés (exon 5-8) [102, 103]. Le gène *TP53* (MIM*191170), localisé sur le chromosome 17 (17p13.1), code pour une phosphoprotéine nucléaire de 53 kDa impliquée dans la réponse au stress cellulaire [104, 105]. *TP53* répond à l'architecture classique d'un facteur transcriptionnel, avec un domaine de transactivation amino-terminal, un large domaine de liaison à l'ADN répondant à des séquences consensus de haute affinité, un domaine d'oligomérisation, trois signaux de localisation nucléaire et un domaine basique participant à la régulation transcriptionnelle en position carboxy-terminale.

En situation **basale**, p53 est présent en faibles quantités, sous une forme instable avec une conformation restreignant ses capacités de liaison de l'ADN, et en localisation cytoplasmique en raison de sa liaison à la protéine Mdm2. Mdm2 régule négativement p53 par trois mécanismes : inhibition de son activité transcriptionnelle par fixation de Mdm2 aux séquences trans-activatrices N-terminales, bloquant ainsi l'interaction avec TFIIH, export nucléaire et ubiquitination via son activité E3 ubiquitine ligase, permettant la dégradation protéosomique de p53. L'expression de *MDM2* est elle-même régulée par p53, constituant ainsi une boucle de rétrocontrôle négatif. La régulation de p53 est également médiée par JNK, qui contribue à sa

dégradation via le protéasome. MDM2 inhibe p53 en phases S et G₂/M, JNK est davantage impliqué en phase G₀/G₁ [106].

En situation de **stress cellulaire** et notamment de dommages à l'ADN [105], la protéine p53 est la cible de diverses modifications post-traductionnelles par les nombreuses kinases impliquées dans la cascade de phosphorylations secondaire à la détection d'un stress cellulaire (ATM, ATR, DNA-PK, Chk1 et Chk2, p38 et JNK). La nature et l'étiologie du dommage induisent une activation spécifique de p53, orientant ainsi la réponse apportée par p53 [106]. La phosphorylation consécutive de p53 induit sa libération de Mdm2 et JNK, et sa stabilisation sous forme tétramérique, mais induit également son interaction avec différents co-facteurs notamment de la famille CBP/p300 et pCAF. Le domaine C-terminal est la cible de modifications post-traductionnelles par acétylation, nécessaire à l'activation de p53 en réduisant son ubiquitination, en re-compartmentalisant p53 dans le noyau et en optimisant sa capacité de liaison à l'ADN. D'autres modifications de p53 telles que phosphorylation, glycosylation, sumoylation sont impliquées. Ainsi activée, p53 agit comme un puissant inducteur transcriptionnel d'un large panel de gènes impliqués dans les mécanismes de contrôle du cycle cellulaire (*GADD45*, *p21WAF1*, *14-3-3*), de réparation de l'ADN (système NER), de l'apoptose (*BAX*, *fas*, *IGF-BP3*) et de la régulation métabolique. p53 s'auto-régule lui-même par un rétrocontrôle négatif sur sa propre transcription [106].

Ces fonctions déterminent ainsi un rôle majeur de *TP53* dans le contrôle cellulaire et la réponse au stress génotoxique. Cette fonction centrale de **suppresseur de tumeur** a conduit à définir *TP53* comme **gardien du génome**. En corollaire de ce rôle central, le gène *TP53* est retrouvé altéré dans un nombre majeur de cancers, et constitue souvent l'un des événements oncogéniques initiaux. On retrouve ainsi des variations somatiques de *TP53* dans plus de 90% des carcinomes séreux de haut grade ovariens [107], 80% des carcinomes basal-like mammaires [108]. Les amplifications de *MDM2* constituent également des événements oncogéniques primordiaux dans certains types tumoraux notamment les liposarcomes, aboutissant de même à une altération de la voie p53.

Altérations constitutionnelles de TP53

A la différence des altérations affectant les gènes suppresseurs de tumeurs, principalement de nature non-sens (variations tronquantes), les variations constitutionnelles répertoriées de *TP53* sont majoritairement des variations faux-sens (73%) [109], réparties sur l'ensemble de la séquence codante mais néanmoins concentrées sur 8 *hotspots* pour environ 30% d'entre elles. Ces variations affectent principalement le domaine de liaison à l'ADN et impactent ainsi l'activité transcriptionnelle. Du point de vue de la pathogénicité, les altérations se distribuent en trois catégories : variations faux-sens **avec effet transdominant négatif**, variations faux-sens **sans effet transdominant négatif** et enfin variations **nulles**. Les variations faux-sens avec effet trans-dominant négatif sont responsables d'une plus grande précocité d'apparition du premier cancer, et sont retrouvées ainsi majoritairement dans les formes pédiatriques du

syndrome de Li-Fraumeni. Ces variants résultent de la tétramérisation d'une forme mutée avec les formes sauvages de p53, aboutissant à un tétramère inactif. Ces variants à effet trans-dominants négatifs n'affectent pas le domaine de liaison à l'ADN ni le domaine d'oligomérisation mais affectent la stabilité structurale de p53 (mutants structurels), produisant des tétramères inactifs [104]. La résultante est donc un déficit cellulaire supérieur de p53 fonctionnel comparativement à la situation d'une variation nulle ou avec perte de fonction qui se limite alors à une haplo-insuffisance. Cette classification fonctionnelle est l'élément déterminant pour interpréter l'hétérogénéité phénotypique au sein du syndrome LFS et proposer une corrélation génotype/phénotype des variations de *TP53*. Les variations faux-sens sont à l'origine d'une plus grande précocité d'apparition du premier cancer comparativement aux variations engendrant une perte de fonction, d'autant plus lorsqu'elles sont comparées aux variations dites nulles [101, 110, 111]. Ainsi, dans la série française de 157 familles rapportée en 2013 [111], l'âge médian d'apparition du premier cancer est de 23 ans en situation de variations faux-sens avec effet dominant négatif versus 45 ans lorsque l'altération résulte d'un réarrangement de grande taille. En utilisant un test fonctionnel basé sur le niveau d'induction transcriptomique des gènes cibles de *TP53* en situation de stress oncogénique (exposition à la Doxorubicine de lignées EBV dérivés de patients et de contrôles), les conséquences fonctionnelles des différentes altérations peuvent être ainsi mesurées. On observe ainsi une diminution modérée du score d'induction en situation de variations dites nulles ou sans effet dominant négatif comparativement à la réduction drastique de ce score en situation de variations faux-sens à effet dominant négatif [111].

Prévalence du cancer du sein dans le syndrome LFS

Dès la description *princeps* du syndrome de Li-Fraumeni, la prévalence du cancer du sein de la femme jeune est notée, puisque 3 des mères d'enfants atteints avaient développé des cancers du sein avant 30 ans [96]. La base de données de l'IARC *TP53* recense, parmi les porteurs de variations délétères constitutionnelles de *TP53*, 29,5% d'individus (exclusivement féminins) affectés par un cancer du sein, versus 12,79% d'individus affectés par un sarcome. Le cancer du sein de la femme jeune constitue ainsi la première pathologie tumorale du syndrome de Li-Fraumeni. Dans les familles françaises porteuses de variations constitutionnelles de *TP53*, le cancer du sein constitue le type tumoral le plus prévalent parmi les adultes de sexe féminin, avec 127 femmes atteintes sur 172 porteuses (79%) et des âges d'apparition précoces (35 ans en moyenne) [101]. Les variations délétères de *TP53* dans la population générale sont rares, estimées à 0,28% dans la cohorte ExaC. L'étude de la contribution de *TP53* dans la survenue de cancers du sein est récente. Plusieurs études [100, 110, 112-120] ont cependant cherché à estimer cette prévalence dans des cohortes de cancers du sein sélectionnés pour la précocité de leur survenue et l'absence de variation détectée des gènes *BRCA1* et *BRCA2* (Tableau 4).

Etudes	Cas index	Nombre de patients	Histoire familiale	Statut BRCA	Variants TP53 (%)
Lalloo et al, 2003 (103)	Cancer du sein <30 ans	82	Sans sélection	négatif	4,9
Walsh et al, 2006 (106)	Cancer du sein <35 ans	21	Agrégation de cancers du sein et de l'ovaire	négatif	0
Bougeard et al, 2008 (107)	Cancer du sein <33 ans	45	Négatif	négatif	6,7
Ginsburg et al, 2009 (108)	Cancer du sein <30 ans	95	Sans sélection	négatif	0
Gonzalez et al, 2009 (109)	Cancer du sein <30 ans	341	Négatif	négatif	7,1
Tinat et al, 2009 (110)	Cancer du sein <36 ans	128	Négatif	négatif	3,9
Mouchawar et al, 2010 (111)	Cancer du sein <30 ans	52	Sans sélection	négatif	3,8
Ruijs et al, 2010 (112)	Cancer du sein <30 ans	12	Négatif	négatif	8,3
McCuaig et al, 2012 (113)	Cancer du sein <30 ans	28	Sans sélection	négatif	21,4 %
Castera et al, 2014 (2018) (114)	Cancer du sein	468	Sans sélection	négatif	1 %
Bakhuizen et al, 2019 (115)	Cancer du sein <30 ans	370	Sans sélection	négatif	2,2 %
Total		1642			3,2 %

Tableau 4 : Contribution des altérations *TP53* dans le cancer du sein précoce

Les cancers du sein dans le contexte d'une prédisposition LFS sont plus fréquemment *HER2* amplifiés [121, 122]. Néanmoins, ce seul critère phénotypique sans argument de précocité ou familial ne permet pas un enrichissement en variants délétères de *TP53*, avec seulement 1,4% de variations délétères de *TP53* identifiée dans une population de cancers du sein *HER2+* de moins de 50 ans [123]. Les carcinomes canaux *in situ* sont une composante de l'expression phénotypique des cancers du sein liés à *TP53* [121, 124].

La prise en charge des patientes avec altération *TP53* repose, selon le protocole de Toronto [125-127] sur la mise en place dès 18 ans d'une auto-palpation puis à partir de 20 ans d'une surveillance sénologique clinique semestrielle, et à partir de 20-25 ans d'une surveillance par IRM mammaire. La mastectomie bilatérale prophylactique peut être, à l'instar de la prédisposition héréditaire liée à *BRCA1/2*, considérée comme une stratégie de réduction de risque. Les récents travaux menés par l'unité Inserm 1245 [111, 128] ont par ailleurs démontré ex

vivo la génotoxicité accrue des traitements par rayonnement ionisant et de la majeure partie des cytotoxiques, conduisant à une prise en compte du statut *TP53* altéré dans les stratégies thérapeutiques.

PALB2

PALB2 (MIM*610355) est un gène du complexe de Fanconi (*FANCN*) situé en 16p12. De par son interaction directe avec *BRCA2*, *PALB2* a fait l'objet d'une approche de type gène candidat avec identification d'altérations parmi des familles de cancers du sein [129-132] et un effet fondateur dans la population Finlandaise [133]. Selon la population étudiée, entre 0,5 et 3% d'altérations de *PALB2* sont ainsi rapportées dans des familles de cancers du sein, principalement des variations pertes de fonction. Les risques relatifs de cancers du sein, la courbe d'incidence selon l'âge et le phénotype des cancers du sein résultant d'une altération de *PALB2* sont très similaires aux données de *BRCA2* [134]. Ainsi, le risque relatif de cancer du sein est estimé à 9,47 avec un risque cumulé de cancer du sein à 70 ans de 47,5%. Les cancers du sein sont majoritairement hormono-sensibles, avec 74% de cancers présentant des récepteurs hormonaux positifs. Le risque de cancer du sein chez les hommes est également significativement augmenté, avec un risque relatif de 8,30. En revanche, à la différence de *BRCA2*, le risque de cancer des ovaires n'apparaît pas augmenté, avec un risque relatif de 2,30 mais non significatif. Tout comme pour *BRCA2*, des altérations délétères de *PALB2* ont été documentées dans des cancers du pancréas mais sans enrichissement démontré, soulevant l'hypothèse d'une contribution aux formes familiales de cancer du pancréas et à un sur-risque pour les individus porteurs. La prise en charge de ces individus porteurs d'altérations de *PALB2* rejoint de ce fait la prise en charge mammaire des porteurs d'altérations *BRCA2*, basée sur la surveillance sérologique renforcée avec IRM mammaire annuelle à partir de 30 ans et la discussion autour de la mastectomie bilatérale prophylactique.

RAD51

RAD51 est une famille de paralogues : *RAD51B* (MIM*602948), *RAD51C* (MIM*602774), *RAD51D* (MIM*602954), *XRCC2* (MIM*600375) et *XRCC3* (MIM*600675). Ces gènes sont des acteurs majeurs de la recombinaison homologue en se fixant sur les monomères d'ADN après activation par *BRCA2*. Les altérations bialléliques de *RAD51C* et *XRCC2* contribuent à la maladie de Fanconi (*FANCO* et *FANCU*). De ce fait, les paralogues de *RAD51* et en premier lieu *RAD51C* ont fait l'objet d'analyses dans le contexte de la prédisposition héréditaire au cancer du sein et de l'ovaire, avec identification de 6 variants délétères monoalléliques de *RAD51C* dès 2010 dans une cohorte allemande de patientes affectées par des cancers du sein et de l'ovaire [135]. Cependant, ces variations étaient systématiquement identifiées dans des familles comportant des cancers de l'ovaire, et absentes dans les agrégations de cancers du sein exclusivement. *RAD51D* de même a

été retrouvé altéré (8/911) dans des cohortes de familles agrégeant cancers du sein et de l'ovaire, et plus particulièrement dans les clusters de cancers de l'ovaire, conférant un *odds ratio* de 6,30 (IC95 2,86–13,85) [136]. La contribution des autres paralogues est plus débattue, et a fait l'objet d'une analyse dédiée dans une cohorte française de 2469 patients adressés pour séquençage sur les critères standards de suspicion de prédisposition héréditaire au cancer du sein et de l'ovaire [137]. Un total de 21 variations différentes est ainsi identifié chez 30 patients, réparties ainsi : *RAD51B* (n = 4), *RAD51C* (n=12), *RAD51D* (n=7), *XRCC2* (n=2) and *XRCC3* (n=5). De nouveau, une prévalence plus importante est identifiée dans les familles affectées par des cancers de l'ovaire (taux de détection de 2,63%) comparativement aux patientes affectées de cancer du sein exclusivement (taux de détection de 0,73%). Considérant la contribution des variations de *RAD51C* et de *RAD51D* dans la prédisposition héréditaire au cancer de l'ovaire [137, 138], ces deux gènes ont été implémentés dans les panels de gènes avec une prise en charge pour les porteurs similaires aux porteurs de variants *BRCA1* ou *BRCA2*, en ce qui concerne la prévention du risque ovarien par l'annexectomie bilatérale prophylactique avant 50 ans et selon l'histoire familiale. Ces altérations sont également considérées comme facteur de sensibilité aux inhibiteurs de PARP [139]. En revanche, concernant le risque mammaire, la contribution de ces rares variants reste débattue et ne justifie pas d'une prise en charge de type femme à haut risque mammaire impliquant l'IRM mammaire ou la mastectomie prophylactique.

Gènes du complexe MRN

Le complexe MRN est un hétérotrimère composé de 2 hétérodimères associant MRE11 et RAD50, associé à une seule molécule NBN (complexe MRE11/RAD50/NBN). Il participe à la réparation de l'ADN par recombinaison homologue, mais aussi à la recombinaison méiotique, le contrôle du cycle cellulaire et le maintien de l'intégrité télomérique. Les altérations bi-alléliques de ces gènes sont à l'origine de prédispositions héréditaires aux cancers par une susceptibilité accrue aux radiations ionisantes, avec pour *MRE11* le syndrome ataxie-télangiectasie-like (*Ataxia-telangiectasia-like disorder-1* ou *ATLD1*, OMIM: 604391), pour *NBN* et *RAD50* le syndrome de Nimègue (*Nijmegen breakage syndrome*, OMIM: 251260 et *Nijmegen breakage syndrome-like* OMIM: 613078). Des altérations hétérozygotes ont été identifiées au sein de ces gènes chez des individus avec présentation évocatrice d'une prédisposition héréditaire au cancer du sein. Sur une cohorte de 2436 individus, un enrichissement en variants rares est ainsi décrit, conférant un *odds ratio* de 2,88 pour les variants évocateurs d'une pathogénicité [140]. Compte tenu de ce sur-risque modeste, ces gènes sont donc reconnus comme des gènes de susceptibilité au cancer du sein de risque intermédiaire.

PTEN

Le syndrome de Cowden (OMIM: 158350) est une maladie de transmission autosomique dominante correspondant à une hamartomatose de localisations cutanée, digestive, mammaire, thyroïdienne et neurologique. Un phénotype évocateur est présent, associant macrocéphalie, trichilemmomas, papules papillomateuses et kératose acrale. Ce syndrome s'associe à une prédisposition aux cancers du sein, de la thyroïde (de type folliculaire) et de l'utérus, ces trois cancers constituant des critères majeurs pour le diagnostic phénotypique [141]. Les bases moléculaires de ce syndrome reposent en grande partie sur des altérations de *PTEN*, gène situé en 10q22-23, de type variants perte de fonction. Selon les critères syndromiques retenus, la contribution des altérations de *PTEN* est autour de 30% [142, 143]. La prévalence des altérations de *PTEN* est faible, avec une estimation de 1/200.000 individus.

L'atteinte mammaire dans le syndrome de Cowden est importante, avec un sur-risque de cancers du sein récemment revu à la hausse, à hauteur du risque lié à *BRCA1* et *BRCA2*. Le risque cumulé à 70 ans est ainsi estimé entre 67 et 85%, débutant à 30 ans et atteignant 50% à 50 ans [144-146]. L'association d'altérations des gènes *SDHx* (*SDHA*, *SDHB*, *SDHC*, *SDHD*) dans les syndromes de Cowden (lié à une altération de *PTEN* ou *PTEN* négatif) apparaît comme un facteur modificateur du risque de cancer, avec une augmentation significative du risque de cancers du sein, du rein et de la thyroïde [147]. La bilatéralité est retrouvée dans 25% des cas [148]. Les recommandations de prise en charge pour le risque mammaire reposent sur une surveillance clinique à partir de 18 ans, et l'initiation d'une surveillance radiologique par mammographie et IRM en cas de densité mammaire forte, à partir de 30 ans ou 5 ans avant le cas le plus précoce dans la famille. La mastectomie prophylactique est souvent discutée dans cette prédisposition, compte tenu des risques cumulés majeurs et de la difficulté de surveillance liée à la plus grande fréquence de mastopathies fibro-kystiques dans cette population [145].

LKB1/STK11

Le syndrome de Peutz-Jegher (OMIM: 175200) est une autre hamartomatose de transmission autosomique dominante, résultant d'altération du gène *LKB1/STK11* situé en 19p13.3. Ce syndrome est marqué essentiellement par une polypose atténuée du tractus digestif, associée à un sur-risque de cancers digestifs. Néanmoins, d'autres sur-incidences de cancers sont décrites dans ces familles, incluant au-delà des cancers gastro-intestinaux les cancers du sein, du pancréas, de l'ovaire, de l'utérus, du testicule et du poumon. Le risque de cancer du sein en particulier est estimé à 29% à 65 ans [149].

CDH1

Les altérations délétères du gène *CDH1* (16q22.1, MIM*192090) sont impliquées dans la prédisposition héréditaire au cancer gastrique (*hereditary diffuse gastric cancer syndrome* ou HDGC-syndrome) (MIM:192090). L'observation de cas fréquents et précoces de cancers du sein dans ces familles, volontiers de type lobulaire, avec un risque cumulé de 20 à 40% [150, 151] ainsi que la fréquente perte d'hétérozygotie associée aux variations constitutionnelles de *CDH1* conduisent à proposer *CDH1* comme gène de susceptibilité au cancer du sein. L'association au sur-risque de cancer gastrique est inconstante, avec une corrélation génotype-phénotype nucléotide-dépendante et mettant en jeu probablement d'autres facteurs modificateurs, conduisant à proposer deux entités syndromiques : le syndrome HDGC et le syndrome HLBC pour *hereditary lobulaire breast cancer*, dénué d'histoire familiale de cancers gastriques [152]. Cette distinction est majeure pour le conseil génétique, puisque la gastrectomie prophylactique avant 20 ans est la recommandation formulée pour les individus affectés par le syndrome HDGC. Selon l'*International Gastric Cancer Linkage Consortium* (IGCLC), les critères de cancers lobulaires du sein bilatéraux avant 50 ans, ou de cancer lobulaire du sein unilatéral avant 45 ans et avec apparentés affectés également par des cancers du sein de type lobulaire sont retenus en l'absence d'histoire évocatrice de cancers gastriques pour l'analyse de *CDH1* [153]. Néanmoins la fréquence des altérations de *CDH1* et leur causalité sont débattues. Dans une méta-analyse regroupant l'analyse de 482 individus affectés par des cancers du sein de type lobulaires et répondant aux critères de l'IGCLC, seules 14 variants de *CDH1* (2,9%) sont rapportés [154] avec des altérations principalement de type faux-sens distribuées sur l'ensemble de la séquence codante. De surcroît, *CDH1* présente un taux de variants de signification inconnue important (rapport variants de signification inconnue sur variants délétères de 13,4) [155] rendant complexe et nécessairement prudente l'intégration de *CDH1* dans les panels de gène analysés pour la prédisposition héréditaire au cancer du sein.

La prise en charge est similaire à celle des individus porteurs de variants *BRCA1* ou *BRCA2*, avec surveillance radiologique basée sur l'IRM mammaire annuelle couplée à la mammographie et à l'échographie, ce d'autant que les foyers lobulaires sont souvent occultes sur la simple mammographie. La chirurgie prophylactique mammaire peut être discutée au cas par cas.

ATM

ATM, localisé sur le chromosome 11q22-23 (MIM*607585), encode une phosphoprotéine impliquée dans la reconnaissance des lésions de l'ADN et l'activation des voies de la réparation (p53, BRCA1). Les altérations homozygotes ou hétérozygotes composites d'*ATM* sont à l'origine d'une maladie génétique rare, l'Ataxie Téléangiectasie (OMIM: 208900), caractérisée par une ataxie

cérébelleuse, des télangiectasies cutanées et oculaires, une immuno-déficiencia et surtout une susceptibilité accrue aux radiations ionisantes et une augmentation du risque de cancers notamment aux hémopathies malignes. La fréquence des hétérozygotes se situe entre 1/100 et 1/200. Les sujets féminins hétérozygotes (vecteurs sains) semblent présenter un risque relatif augmenté de cancers du sein, avec une estimation réévaluée à la baisse ces dernières années pour se situer autour de 2,3 [156, 157]. A l'instar de *TP53* a été évoquée une possible modulation du risque de cancer selon la nature faux sens ou pertes de fonction des variants, avec un effet dominant négatif pour les altérations faux sens [158].

CHEK2

Le gène *CHEK2* situé en 22q12.1 (MIM*604373) code pour une protéine kinase impliquée dans le contrôle G2 du cycle cellulaire et de ce fait, impliquée également dans la réparation de l'ADN. La description du variant *CHEK2*1100delC*, délétère car abolissant la fonction kinase de la protéine, dans une famille LFS *TP53* négative [159] a conduit initialement à proposer *CHEK2* comme gène majeur de la prédisposition héréditaire au cancer du sein. Les données en population saine ont néanmoins révélé une fréquence allélique de ce variant relativement élevée autour de 1% et un sur-risque modéré autour de 2 pour les porteurs [160]. Un enrichissement de ce variant est noté dans les familles agrégeant de multiples cancers du sein, avec cancers du sein bilatéral ou cancer du sein masculin, sans argument de plus grande précocité en revanche [161, 162]. Le variant *CHEK2*1100delC* pourrait ainsi, de par sa fréquence et son risque relatif modéré, contribuer à l'explication d'agrégaions familiales de cancers du sein (*familial cluster*), et/ou agir en tant que facteur modificateur d'un autre gène de susceptibilité. Cependant, ce variant ne modifie pas le risque lié à *BRCA1* ou *BRCA2* en situation de co-occurrence [160]. L'intérêt de la recherche de ce variant est donc limité, eu égard à l'augmentation modérée du risque mammaire et de la probable interaction de ce variant avec d'autres gènes de susceptibilité, dans un déterminisme oligogénique à ce jour non appréhendé.

Autres gènes impliqués ou suspectés

Concernant le risque ovarien spécifiquement, *BRIP1* et les gènes MMR (*MLH1*, *MSH2*, *MSH6* *PMS2*) doivent être également considérés.

BRIP1 (MIM*605882) est un gène du complexe Fanconi (*FANCD1*) et interagit avec *BRCA1* lors de la recombinaison homologue. Des altérations délétères sont rapportées dans des cohortes familiales de cancers du sein, avec des taux de variations inférieurs à 1% et des *odds ratio* faibles ou non significatifs pour le risque de cancers du sein [163-165]. Par l'analyse de cohortes constituées de familles présentant des cancers du sein exclusivement, association de cancers du

sein et de l'ovaire ou cancers de l'ovaire exclusifs, *BRIP1* apparaît comme un gène de prédisposition au cancer de l'ovaire exclusivement, avec une prévalence de 2,55% et des *odds ratios* > 20 lorsque sont considérés uniquement les individus de la cohorte cancer de l'ovaire exclusif [166]. Les âges d'apparition de ces néoplasies ovariennes apparaissent plus tardifs, au-delà de 61 ans. En comparaison, le taux de variations de *BRIP1* dans les cohortes de cancers du sein et de l'ovaire ou de cancers du sein exclusifs s'avèrent inférieur à 1%, avec des *odds ratios* non significatifs. Le risque cumulé à 80 ans de survenue d'un cancer de l'ovaire en présence d'altérations de *BRIP1* est estimé à 5,8%, conduisant à discuter de stratégies de prévention du risque par annexectomie bilatérale prophylactique.

BARD1, situé en 2q34-35 (MIM*601593), interagit avec *BRCA1* mais également p53, et contribue ainsi à la réparation de l'ADN. L'identification de variants de *BARD1* dans des familles évocatrices de prédisposition héréditaire au cancer du sein ou de l'ovaire [167], notamment le variant Cys557Ser [168] ou Glu564X [169], conduit à évoquer *BARD1* comme gène mineur de prédisposition, avec des effets fondateurs suspectés et un risque relatif de cancer du sein notamment autour de 5,8 dans la population finlandaise et 2,3 dans la population polonaise respectivement. Néanmoins l'absence de co-ségrégation, l'absence de perte d'hétérozygotie dans les tumeurs, la fréquence de ces altérations dans les populations contrôles rendent prudents l'interprétation de ces variations de *BARD1*. Proposé comme contribuant à un modèle polygénique, *BARD1* Cys557Ser (rs28997576) ne semble pas en revanche agir comme gène modificateur chez les porteurs de variants *BRCA1* ou *BRCA2*, en raison de l'absence de co-occurrence [170]. Néanmoins dans la plus large série allemande, où l'absence d'implication de *BARD1* dans la prédisposition au cancer de l'ovaire est confirmée, un *odds ratio* de 5,35 significatif est en revanche identifié pour la prédisposition au cancer du sein, avec un enrichissement dans les formes familiales de cancers du sein de plus grande précocité [171]. Les variations de *BARD1* sont également retrouvées de manière significative dans des cohortes de cancers du sein triple négative [172].

RINT1 (*RAD50-interacting protein 1*) a été analysé de manière extensive dans une population évocatrice de prédisposition héréditaire au cancer du sein, du fait de son implication dans la voie de la réparation de l'ADN [173]. L'identification de variants rares de *RINT1* avec des *odds ratios* de 3,24 conduit à proposer *RINT1* comme gène mineur de prédisposition au cancer du sein. Néanmoins, en l'absence de reproduction de ces résultats dans une seconde étude cas-témoin [174], la contribution de *RINT1* reste débattue.

Parmi les gènes du complexe Fanconi, d'autres candidats ont été proposés. Ainsi, une variation non-sens de *FANCM* (c.5791C>T; rs144567652) a été rapportée présente de manière enrichie dans une population de cancers du sein familiaux comparativement à une population contrôle (OR 3,93) [175]. *SLX4/FANCP* a été également au même titre séquencé dans différentes populations de cancers du sein précoces ou familiaux avec des données contradictoires [176-179].

RECQL encode une *DNA Helicase* impliquée dans la réparation, la réplication, la recombinaison et la transcription de l'ADN. L'identification de variants pertes de fonction ou faux-sens de *RECQL* dans différentes populations [180-182] avec des effets fondateurs suspectés (Pologne, Chine) conduisent à considérer *RECQL* comme potentiel gène de susceptibilité au cancer du sein. Cependant, les *odds ratios* faibles et les études contradictoires dans des populations similaires [183-186] invitent une fois de plus à la plus grande prudence dans la reconnaissance d'une contribution de ces variants à une prédisposition héréditaire au cancer du sein.

Les gènes MMR constituent le *substratum* moléculaire du syndrome HNPCC (*Hereditary Non-Polyposis Colorectal Cancer*) ou syndrome de Lynch, avec les gènes : *MSH2* (MIM*609309), *MSH6* (MIM*600678), *PMS2* (MIM*600259), et *MLH1* (MIM*120436). Les cancers colorectaux et les adénocarcinomes de l'endomètre composent la majeure partie du spectre tumoral, avec néanmoins dans le spectre élargi une augmentation de l'incidence des cancers de l'ovaire [187]. Une méta-analyse décrit un type histologique préférentiellement de type mixte (mucineux/endométrioïde/cellules claires) ou endométrioïde pur, un âge d'apparition précoce (âge moyen de 45 ans), une majorité de stade précoce FIGO I/II (65%) associée à une survie plus favorable (80% à 5 ans), une prédominance des variants *MLH1* et *MSH2* (respectivement 38 et 47%) [188]. Les risques cumulés de cancer de l'ovaire à 70 ans pour les porteurs de variations *MLH1*, *MSH2* et *MSH6* sont respectivement de 20%, 24% et 1% [189]. Les âges d'apparition sont néanmoins postérieurs à 50 ans, légitimant une hystérectomie non conservatrice à 50 ans (ou 5 ans avant l'âge d'apparition le plus précoce dans la famille) pour la prévention du cancer de l'endomètre et de l'ovaire chez les femmes porteuses. L'extinction immuno-histochimique ou l'instabilité microsatellitaire (MSI) constituent des marqueurs phénotypiques du syndrome HNPCC pouvant être recherchés sur le tissu tumoral afin d'orienter la pertinence et le choix du gène MMR à séquencer. Dans une série de cancers de l'ovaire sans sélection sur l'âge ou l'histoire familiale, mais présentant un sous-type à cellules claires [190] ou endométrioïde [191], une extinction des gènes MMR était retrouvée dans les 2 cas autour de 6%, après exclusion des extinctions *MLH1*/*PMS2* toutes liées à une méthylation du promoteur de *MLH1*. Une recommandation de dépistage des tumeurs ovariennes de sous-type à cellules claires ou endométrioïde est ainsi proposée par le Collège Américain de Pathologistes. A l'instar des inhibiteurs de PARP dans le traitement des cancers de l'ovaire avec déficit dans la voie de la recombinaison homologue, les inhibiteurs de *checkpoint immunitaire* (anti-PD-1/PDL-1) constituent une arme thérapeutique majeure dans les cancers de l'ovaire avec instabilité microsatellitaire dans le contexte HNPCC [192].

L'hypothèse d'une contribution des variants des gènes MMR dans la prédisposition héréditaire au cancer du sein est alimentée par diverses observations, telle que l'incidence accrue de cancers du sein dans des cohortes de patientes atteintes de syndrome de Lynch, avec une prédominance de variants des gènes *MSH2* et un risque relatif de cancer du sein pour les porteuses de variants *MSH2* estimé à 3,11 [193]. Un autre registre de patientes atteintes de cancer de l'endomètre dans un contexte HNPCC rapporte un risque relatif de cancer du sein de

2,51 [194]. La mise en évidence d'un statut MSI dans les tumeurs mammaires des patientes affectées par un syndrome HNPCC contribue à l'hypothèse d'une implication des gènes MMR dans la carcinogenèse mammaire de ces patientes [195], bien que cela soit controversé par une large série documentant le phénotype MMR dans les tumeurs secondaires de patients HNPCC, incluant les tumeurs mammaires et sans statut MMR déficient retrouvé dans ces tumeurs secondaires [196]. Un enrichissement en variants rares de *MSH6* est également rapporté dans une population de familles hollandaises associant soit une agrégation de cancers du sein et de cancers colorectaux, soit plus restreintes à des agrégations de cancers du sein [197]. Bien que la contribution des altérations des gènes MMR ne puisse être exclue dans des formes familiales de cancers du sein, soulevant l'hypothèse d'une extension phénotypique, ces données apparaissent néanmoins insuffisantes, avec études contradictoires et probables facteurs confondants [198].

Le tableau 5 propose une synthèse des implications de ces différents gènes dans la prédisposition au cancer du sein et de l'ovaire, basée sur la publication de Castera *et al* [199].

Gène	P _{patient}	P _{Exac}	Odds ratio
<i>ATM</i>	1,04	0,33	3,20 (2,14 - 4,53)
<i>BARD1</i>	0,19	0,10	2,00 (0,74 - 4,10)
<i>BRCA1</i>	3,70	0,29	13,22 (10,01 - 17,22)
<i>BRCA2</i>	3,95	0,48	8,61 (6,78 - 10,82)
<i>BRIP1</i>	0,48	0,20	2,49 (1,42 - 3,97)
<i>CDH1</i>	0,05	0,011	6,82 (0,49 - 28,32)
<i>CHEK2</i>	1,14	0,69	1,67 (1,18 - 2,27)
<i>MLH1</i>	0,23	0,15	1,59 (0,71 - 2,91)
<i>MRE11A</i>	0,16	0,095	1,74 (0,63 - 3,57)
<i>MSH2</i>	0,29	0,23	1,34 (0,56 - 2,62)
<i>MSH6</i>	0,25	0,10	2,50 (1,12 - 4,67)
<i>NBN</i>	0,22	0,17	1,35 (0,54 - 2,64)
<i>PALB2</i>	0,89	0,11	8,22 (4,91 - 13,05)
<i>RAD51B</i>	0,046	0,044	1,12 (0,12 - 3,46)
<i>RAD51C</i>	0,53	0,12	4,54 (2,55 - 7,48)
<i>RINT1</i>	0,13	0,14	0,90 (0,18 - 2,26)
<i>TP53</i>	0,43	0,28	1,56 (0,85 - 2,52)

Tableau 5 : Prévalence et *odds ratios* des principaux gènes candidats dans la prédisposition

Prédisposition héréditaire au cancer du sein et de l'ovaire - Approches par panel multi-gènes et contributions relatives

L'avènement du séquençage à haut débit (*Next-Generation Sequencing* ou NGS) et l'effondrement des coûts de cette technique ont permis d'accéder au génotype de milliers de patients, dans des approches allant du séquençage à forte profondeur d'un gène unique à des approches exomiques (*Whole Exome Sequencing* ou WES) ou génomiques (*Whole Genome Sequencing* ou WGS). La question de la sélection des patients analysés demeure cruciale, ce d'autant que l'analyse de larges séries est désormais accessible. La question de l'interprétation est encore davantage critique, compte tenu de la mise à disposition d'une quantité massive d'informations. Les limites technologiques et intellectuelles qu'offrent le séquençage de nouvelle génération ont pu conduire à une surestimation et à une interprétation erronée des variants identifiés par cette méthode. Ainsi, la comparaison de 9 génomes d'individus d'ascendance européenne à un génome de référence conduit à 3,8 millions de variants par individu, dont 0,6 millions de variants rares [200]. Au moins 2 variants par individu séquencé apparaissent être rares ou non répertoriés, prédits délétères et impliqués dans le déterminisme d'une pathologie, leur conférant le statut de très probable variant délétère. Néanmoins la plupart de ces variants identifiés sont des faux positifs, résultant d'erreurs d'annotation ou d'erreurs de séquençage. Des efforts de curation sont donc essentiels, compte tenu des disparités inter-laboratoires observées. Ainsi, dans un registre prospectif, 26% des variants rapportés par deux laboratoires différents sont interprétés de manière conflictuelle [201]. L'ACMG-AMP (*American College of Medical Genetics and Genomics - Association of Molecular Pathology*) a contribué fortement à une uniformisation et amélioration des principes régissant l'interprétation et l'annotation de ces variants [22]. Une uniformisation de la nomenclature est ainsi imposée aux Laboratoires, mais surtout une démarche pondérée d'interprétation est définie, pour aboutir aux 5 classes de variants définies par Plon *et al* [202] : (i) Classe 1 : bénin, (ii) Classe 2 : probablement bénin, (iii) signification indéterminée, (iv) probablement pathogène, (v) pathogène. Différents items pondérés régissent cette classification, comprenant : nature de la variation (non-sens, indels avec ou sans décalage du cadre de lecture, faux-sens ou synonyme, intronique), prédictions bio informatiques de pathogénicité au travers de l'impact sur l'ARN ou la protéine, données de ségrégation, données d'occurrence dans une population contrôle, variant répertorié dans une base de données clinico-biologiques, variants *de novo*, données fonctionnelles, données alléliques (variant pathogène en *cis*). Par ailleurs, au-delà de ce cadre générique, des démarches d'interprétation gène-spécifique demeurent nécessaires, telles que présentées pour *BRCA1/2* par le groupe ENIGMA [203] ou pour les gènes *MMR* par le groupe inSiGHT [204]. En dépit de cette standardisation dans l'interprétation, des discordances inter-laboratoires subsistent dans une large proportion, de l'ordre de 22% pour les différences de classification avec impact clinique [29]. Par ailleurs, au-delà de l'interprétation, la communication de ce résultat aux cliniciens et aux patients doit être effectuée de manière pédagogique, le vécu familial et personnel à l'origine de la démarche d'analyse entraînant un biais d'interprétation supplémentaire. Ainsi dans une

population américaine, sur une période récente (2014-2015), plus de 50% des patientes affectées par un cancer du sein localisé dans un contexte familial effectuaient une mastectomie bilatérale prophylactique sur la base d'un variant de classe 3 [205]. A l'inverse, une possible sous-estimation de l'implication de variants autres que ceux de *BRCA1* et *BRCA2* peut être invoquée, les filtrations de variants s'appuyant principalement sur la recherche de variations tronquantes (non-sens, altération des sites canoniques d'épissage) à l'instar des altérations de *BRCA1* et *BRCA2* et d'autres gènes suppresseurs de tumeurs. L'interprétation des variants faux-sens, plus délicate, s'appuie sur des données de conservation nucléotidique ou protéiques, sur des logiciels de prédiction bio informatiques et sur d'éventuelles données fonctionnelles. Une cohorte de 1297 patientes affectées de cancers du sein avant 45 ans a été ainsi analysées pour 9 gènes hors *BRCA1* et *BRCA2* : *ATM*, *CHEK2*, *MRE11A*, *RAD50*, *NBN*, *RAD51*, *RINT1*, *XRCC2* et *BARD1* [206]. La filtration bioinformatique conservait les variants rares (MAF <0,1%) et utilisait les données de 4 logiciels de prédiction de pathogénicité pour les variants faux-sens. Les auteurs ont obtenu des *odds ratio* > 2,5 pour des variants rares affectant ces gènes, comparativement à une population contrôle, en conservant les variants rares affectant les domaines fonctionnels importants de ces protéines mais aussi en proposant un modèle polygénique.

Un des éléments clés de cette démarche interprétative est l'occurrence du variant identifié dans une population contrôle. Les projets génomiques en populations contrôles, indemnes de pathologies de déterminisme génétique, sans néanmoins que le processus de sélection de ces contrôles ne permette d'exclure formellement tout phénotype évocateur d'une prédisposition héréditaire, ont révolutionné la vision du généticien en rapportant une fréquence insoupçonnée de variants rares (inférieur à 0,1%) dans ces populations. Ainsi, dans une étude australienne comportant 2000 patients affectés de cancers du sein ou de l'ovaire avec une présentation personnelle et/ou familiale justifiant d'une analyse génétique, et une population contrôle de 1997 sujets indemnes de cancers, des variations non-sens ou faux-sens sont largement retrouvées dans la population contrôle, parfois dans des proportions supérieures à la population affectée [207]. La valeur prédictive positive d'une variation identifiée n'est de ce fait que de 0,71, soulignant la précaution requise dans l'interprétation de ces variations et leur imputabilité. La cohorte ExAC est constituée de plus de 60000 individus présumés indemnes et reflétant la diversité ethnique de la population, dont les exomes sont publiques [208]. On retrouve dans cette base de données génomiques presque 3000 variants parmi 19 gènes retenus pour leur implication dans des pathologies de déterminisme Mendélien, dont *BRCA1* et *BRCA2*. Une fraction importante (n=868) de ces variants de classe 4 ou 5 selon la classification ACMG/AMP sont retrouvés dans la population ExAC à des fréquences alléliques inférieures à la fréquence allélique définie pour chaque gène comme supportant une pathogénicité certaine (*Maximal Pathogenic Allele Frequency* ou MPAF), confirmant ainsi que la population ExAC n'est pas enrichie en individus porteurs de variants pathogènes. En revanche, ces données génomiques illustrent le polymorphisme majeur des populations, avec un nombre important de variants rares (MAF<0,1%). Pour *BRCA1* et *BRCA2*, 86 variants excédant la MPAF sont rapportés, la totalité non pathogènes, soit de classe 1/2 (98% des variants), soit de classe 3.

Les analyses génétiques par panel de gènes offrent l'avantage, à coût quasi constant, d'analyser de manière synchrone plusieurs gènes et ainsi éviter des analyses séquentielles consommatrices de ressources et de temps. On trouve ainsi référencés, dans le champ de la prédisposition héréditaire au cancer du sein, de l'ovaire et de la prostate, plus de 200 panels commerciaux ou académiques sur le site du NCBI [209]. Ces panels sont composés de quelques dizaines à plusieurs centaines de gènes, avec séquençages complets ou uniquement des séquences codantes, ou encore de *hot spots* d'intérêt, et construits essentiellement autour des gènes de la réparation de l'ADN. De nombreuses publications [119, 165, 199, 210-216] rendent compte de ces approches multigènes, avec des populations analysées et des méthodes d'interprétations diverses, complexifiant la compréhension des bases moléculaires et la contribution réelle des variations identifiées (Tableau 6).

Les tailles des séries analysées, les critères de sélection des individus, la définition des panels rendent compte des différences observées. Cependant, on constate une relative constance dans les distributions des variations observées, avec environ 10% des individus analysés présentant une altération, et autour de 50% de ces altérations concernant *BRCA1* ou *BRCA2*. Les variants d'*ATM*, *PALB2* et de *CHEK2* viennent ensuite par ordre de fréquence. L'implication de *TP53* est variable, fonction de la population étudiée et notamment du critère d'âge. L'analyse conjointe d'une population contrôle, ou l'extraction des bases de données de population générale (ExAC) permet dans ces séries d'apprécier les *odds ratio* attribuables à ces altérations. On constate, du fait d'une tentative de réduction de l'hétérogénéité génétique de la population contrôle dans l'étude de Castera *et al.* [199] par des témoins de même origine géographique, et séquencés conjointement avec des critères de filtration similaire, que les *odds ratios* deviennent fortement significatifs, en comparaison de l'étude plus large mais avec population contrôle issue des données ExAC proposée par Couch *et al.* [213]. Ainsi, au-delà des *odds ratio* élevés pour *BRCA1* et *BRCA2* dans la population HBOC, Castera *et al* observent des résultats également significatifs pour *PALB2* dans les familles avec cancers du sein exclusivement et avec un *odds ratio* de même importance que pour *BRCA2*, mais également des *odds ratios* significatifs uniquement dans les familles comportant des cancers de l'ovaire pour *RAD51C*, *RAD51D* et *BRIP1*, et enfin significatifs uniquement dans les familles présentant un cancer du sein de type triple négatif pour *BARD1*.

Etudes	Ref.	Population	N	Panel	Méthode	Résultats
Buyt et al, Cancer 2017	(197)	Cancer du sein <45 ans ou FH+	35000	25 gènes (Myriad MyRisk)	Exons et RGT Classes 4 - 5	9,3% ≥ 1 variant Distribution (%) : BRCA1 24 - BRCA2 24,4 - CHEK2 11,7 - ATM 9,7 - PALB2 9,3 - BARD1 2 - BRIP1 3,2 - TP53 1,8 - gènes MMR 7
Castera et al, EJHG 2014	(119)	Critères minimaux français pour analyse BRCA1/2	708	27 gènes	Exons et RGT Classes 4 - 5	14% ≥ 1 variant Distribution (%) : BRCA1 30,4 - BRCA2 28,6 - PALB2 6,2 - TP53 5,4 - ATM 4,5 - CHEK2 4,5 - MSH2 4 - PMS2 2,7 - MRE11A 2,6 - NBS1 2,6 - RAD51C 2,6 - CDH1 <1
Couch et al, JAMA Onco. 2017	(213)	Patientes analysées par test commercial pour cancer du sein - 59,7% < 50 ans	65057	21 gènes	Population contrôle ExAC	10,2% ≥ 1 variant (6,18% hors BRCA1 et BRCA2) Distribution (%) hors BRCA1/2 : CHEK2 1,73 - ATM 1,06 - PALB2 0,87 Variants associés à haut risque de cancer du sein : BRCA - CDH1 - PTEN - TP53 - PALB2 Variants associés à risque modéré de cancer du sein : ATM - BARD1 - CHEK2 - RAD51D
Couch et al, JCO 2015	(210)	Cancer du sein triple négatif sans sélection sur l'âge ou FH	1824	17 gènes		14,6% ≥ 1 variant Distribution (%) : BRCA1 8,5 - BRCA2 2,7 - PALB2 1,2 - autres <1% Enrichissement chez les individus plus jeunes ou avec FH
Ellingson et al, BCRT 2015	(211)	Cancers du sein avec indication de CTNA	124	142 gènes	WES avec analyse in silico de 142 gènes Classes 4-5	21 % ≥ 1 variant Distribution (%) : BRCA1 3,2 - BRCA2 6,5 - ATM 2,4 - MUTYH 1,6 - autres <1%
Girard et al, IJC 2019	(216)	Cohorte GENESIS Cancers du sein BRCA négatif avec soeur atteinte	1721	113 gènes	Exons et RGT Classes 4-5 Contrôles issus de l'entourage	Sur-risque de cancer du sein uniquement pour PALB2, ATM, CHEK2 et MAST1
Hauke et al, Cancer Med. 2018	(215)	Cancers du sein avec critères minimaux de GC-HBOC et BRCA négatifs	5589	8	Pas de données RGT Classes 4-5	6,1 % ≥ 1 variant (4,9% si restreint aux PTVs) Distribution (%) : ATM 1,27 - CHEK2 1,84 - PALB2 1,15 - autres <1% ORs significatifs pour ATM CHEK2, PALB2, CDH1 et TP53.
Kraus et al, IJC 2017	(214)	Cancers du sein avec critères minimaux de GC-HBOC	581	14	Exons et RGT Classes 4-5	18 % ≥ 1 variant Distribution (%) : BRCA1/2 12,4 - CHEK2 1,7 - PALB2 1,0 - NBN 1,0 - RAD51C 0,9 - ATM 0,5 - TP53 0,3
Li et al, JMG 2016	(212)	kConFab BRCA1/2 négatifs	684	19	Exons et RGT Classes 4-5 Analyses de co-ségrégation	<2% ≥ 1 variant actionnable hors BRCA1/2 Distribution (%) : BRCA1 1,9 - BRCA2 1,6 - CHEK2 2 - PALB2 2 - ATM 1,2% - Autres <1% dont TP53 0,8
Castera et al, Genetics Med. 2018	(199)	Critères minimaux français pour analyse BRCA1/2	5131	34	Exons et RGT Classes 4-5 Biais statistiques corrigés Population contrôle FREX avec correction de l'hétérogénéité génétique	Distribution (%) : BRCA1 3,9 - BRCA2 3,7 - CHEK2 1,1 - ATM 1,0 - PALB2 0,9 - RAD51C 0,5 - TP53 0,5 BRIP1 0,5 Variants associés à haut risque de cancer du sein : BRCA1/2, PALB2, RAD51C/D Variants associés à risque modéré de cancer du sein : ATM, BRIP1, CHEK2, FAM175A, MLH3, PMS1

Tableau 6 : Principales études analysant par panels de gènes des cohortes de cancers du sein

Le paysage moléculaire de la prédisposition héréditaire au cancer du sein et de l'ovaire peut ainsi être considéré comme multiple, avec principalement *BRCA1* et *BRCA2*, mais pour les 50% restants une myriade d'altérations impactant d'autres gènes à des fréquences allant de 7% à moins de 1% (Figure 3) [199].

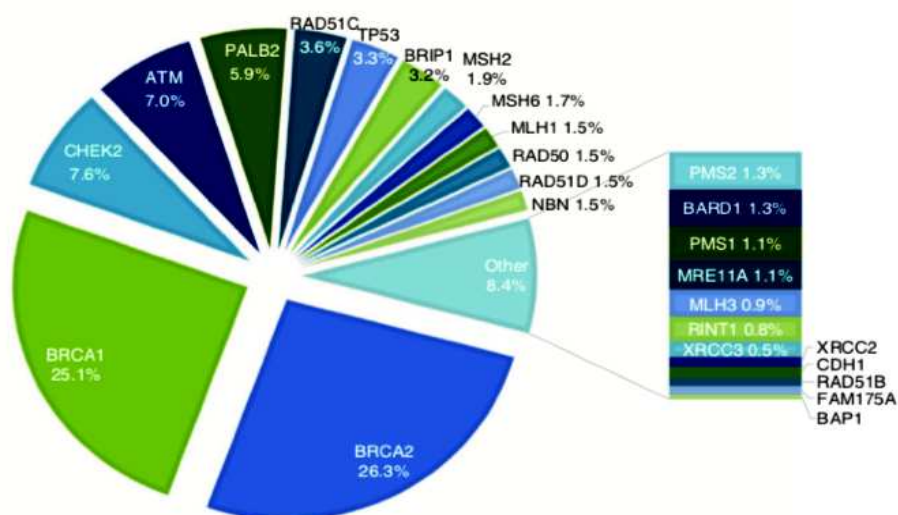


Figure 3 : Paysage moléculaire des cancers du sein héréditaires. Castera *et al*, Genetics in Medicine 2017 [199]

Variations privées - Phénotypes extrêmes

Au-delà de ces altérations génétiques contribuant à un haut risque ou à un risque intermédiaire de cancer du sein et/ou de l'ovaire, une large fraction de familles évocatrices de prédisposition héréditaire demeure orpheline de diagnostic moléculaire. Cette fraction inconnue (*missing hereditary*) est difficilement estimée entre 50 à 75% des familles évocatrices, selon notamment les critères retenus pour suspecter une forme familiale de cancer du sein. Ainsi, avec les critères minimaux retenus pour analyse génétique dans la série de patients du nord-ouest français [199], et un panel élargi comportant l'ensemble des gènes canoniques de la prédisposition HBOC mais aussi d'autres gènes putatifs, 647 variants de classe 4/5 sont identifiés chez 4409 individus soit 85% de patients-familles sans diagnostic moléculaire. Dans une autre revue, cette fraction inconnue est estimée à 51% à la lumière de l'ensemble des contributions moléculaires identifiées et de leurs contributions relatives [217].

Le **reclassement** de variants de classe 3 en variants délétères (classe 4 ou 5) pour *BRCA1* et *BRCA2*, mais aussi pour d'autres gènes de susceptibilité, contribuera certainement à augmenter la fraction expliquée de cette prédisposition. Ainsi, des variants introniques profonds et des variants exoniques faux-sens peuvent s'avérer délétères en altérant l'épissage de *BRCA1* ou *BRCA2*. L'identification de ces effets sur l'épissage permet ainsi leur reclassement en variants pathogènes. Une variation intronique profonde de *BRCA2* (c.6937+594T > G) est ainsi rapportée

chez 9 familles comme associée à une insertion d'exon cryptique de 95 acides aminés entre les exons 12 et 13, par activation d'un site donneur cryptique [218]. Néanmoins, l'étude systématique de l'ARN (RNA-Seq) en complément du séquençage complet de *BRCA1/2*, à la recherche de ces effets sur l'épissage offre néanmoins une très faible contribution [219]. D'éventuelles variations affectant les domaines 5' ou 3'-UTR pourraient également contribuer au risque héréditaire de cancers du sein ou de l'ovaire. Ainsi, la variation en c.-107A>T de *BRCA1* rapportée dans 2 familles non reliées est pathogène en aboutissant à une hyperméthylation du promoteur de *BRCA1* et à une abolition complète de l'expression de l'allèle muté [220]. Des variations survenant en dehors de la séquence codante peuvent également affecter la transcription ou la traduction, en perturbant le fonctionnement des séquences régulatrices de la région promotrice, les séquences *IRES* ou le signal de polyadénylation, ou encore les séquences reconnues par des microARN participant à la régulation de l'expression. Le séquençage de la région 3'UTR de *BRCA1* chez 70 familles évocatrices de prédisposition héréditaire au cancer du sein et de l'ovaire, avec analyse *BRCA1/2* négative, identifiait ainsi 2 variants rares mais sans impact fonctionnel néanmoins dans un système rapporteur impliquant la luciférase [221]. Dans une autre série recherchant de manière similaire des variants 3'UTR de *BRCA1*, la variation c.*1340_1342delTGT est associée à une réduction d'expression de 25% dans le système luciférase lorsque co-exprimé avec le miRNA miR-103 [222]. Les altérations peuvent également impliquer les microARN impliqués dans la régulation de gènes suppresseur de tumeurs, et ainsi contribuer à une prédisposition héréditaire. Ainsi, une délétion de 6 paires de bases dans le cluster miR-17-92 est rapportée dans 2 familles chiliennes avec histoire familiale de cancer du sein et absente dans une population contrôle, avec probable perte fonctionnelle [223].

Les approches par études d'association pangénomique (*Genome-wide Association Study* ou *GWAS*) offrent de potentiels candidats géniques pour cette fraction inconnue. Ainsi, dans une méta-analyse regroupant 9 études de type *GWAS*, permettant l'analyse des *SNPs* de plus de 10.000 patients atteints de cancers du sein et de plus de 12.000 contrôles, 41 nouveaux loci sont identifiés comme significativement associés au risque de cancer du sein [224]. Les *odds ratios* associés à ces loci sont cependant faibles, de l'ordre de 1,10 à 1,20. Par ailleurs, l'interprétation de ces études *GWAS* est très débattue, avec des reproductibilités souvent insuffisantes, des biais statistiques considérables et une hétérogénéité obligatoire de la population étudiée. En particulier, les individus atteints dans cette méta-analyse sont issus d'études multiples dont le critère d'inclusion est parfois uniquement l'existence d'un cancer du sein sans notion de précocité ou d'histoire familiale. Néanmoins, l'identification par ce biais de nouveaux *loci* associés à des surrisques faibles contribue à proposer un modèle polygénique, similaire aux modèles développés notamment dans les troubles neuro-psychiatriques tels que la schizophrénie [225] ou les troubles autistiques [226]. A ces risques génétiques faibles se combineraient des facteurs environnementaux ou de modes de vie, expliquant ainsi les pénétrances faibles et des présentations individuelles ou familiales modérées. Ainsi, une combinaison de 313 *SNPs* issus d'études *GWAS* conduit à proposer un score de risque polygénique associé à un risque de

développer un cancer du sein à 80 ans de 32,6% pour le score le plus élevé [227]. Ce score, inclus dans les modèles prédictifs de risque tel que le BODICEA, permet de moduler significativement le risque prédit en association aux autres paramètres du modèle que sont l'existence de variations génétiques, l'histoire familiale, l'anamnèse des facteurs de risque et la densité mammaire [74].

La recherche de variants rares hérités par séquençage exomique dans des familles évocatrices de prédisposition héréditaire conduit à proposer un modèle de déterminisme Mendélien, monogénique, basé sur des **variants privés**. La relecture de la variabilité du génome humain, grâce au séquençage de nouvelle génération, a ainsi révélé que chaque exome contenait en moyenne plus de 20.000 variants dont 1500 correspondant à des variants rares (MAF<1%) et de 100 à 800 d'entre eux prédits délétères [228]. Ainsi, une analyse exomique de 3 familles évocatrices de prédisposition héréditaire au cancer du sein a conduit à identifier chez les cas index des variants rares prédits délétères affectant notamment les voies de la réparation de l'ADN [229]. Ces variants rares étaient, comme attendu, partagés au sein de chaque famille, mais différents d'une famille à l'autre, introduisant ainsi ce concept de variation privée. L'annotation KEGG de ces variants retrouve en outre des enrichissements différents en termes de voies biologiques entre les 3 familles. De manière parallèle, une large étude saoudienne de séquençage exomique ou par panels de gènes élargis comportant 1000 cas index affectés par des anomalies du développement, troubles neurologiques ou psychiatriques, a identifié dans 34% des altérations de classe 4 et 5 [230]. L'interprétation de ces variants, délicate, trouvait néanmoins une cohérence sur le plan physiopathologique avec le phénotype observé chez les cas index conduisant à proposer 75 nouveaux gènes candidats. 67% des variants identifiés étaient des variants privés. Néanmoins un biais important dans cette étude était l'importante consanguinité présente dans cette population Saoudienne, générant une homozygotie fréquente pour des variants délétères à l'origine de pathologies autosomes récessives, voire des génotypes homozygotes pour des *loci* impliqués dans des pathologies répondant à un mécanisme dominant. Ce biais de consanguinité permet cependant d'appréhender, dans une situation génétique particulière et exceptionnelle, de nombreux gènes candidats pour des pathologies génétiques récessives ou dominantes.

Au-delà de ces approches recherchant des altérations héritées, émerge le **concept de variant de novo**. L'accessibilité au génome offerte par le séquençage de nouvelle génération a permis d'appréhender l'extrême mutabilité du génome humain, et ainsi d'observer à chaque génération l'apparition de variations délétères pouvant contribuer à l'apparition de pathologies, notamment néoplasiques. Ces variants *de novo* surviennent lors de la gamétogénèse, avec une probabilité de $1,18 \cdot 10^{-8}$ par position nucléotidique et par génome haploïde, soient environ 74 variations ponctuelles (SNVs) par génération (Figure 4) [231, 232]. Cette mutabilité est position-dépendante, influencée notamment par la richesse en îlots CpG ou affectant préférentiellement les régions riches en GC et plus fréquemment de type transitionnelle. Des clusters de variants *de novo* sont ainsi décrits, traduisant un environnement génétique particulier propice à une plus grande mutabilité [233]. Ces variants *de novo* sont aussi dépendants de l'origine parentale avec

une origine paternelle préférentielle et une influence de l'âge paternel lors de la conception [234]. Le ratio de 2 à 7 observé entre les variants *de novo* d'origine maternelle ou paternelle s'expliquent par les différences fondamentales dans les gamétogénèses mâles et femelles, avec pour les femmes une phase proliférative mitotique achevée à la naissance, l'activité méiotique seule persistant après la naissance et ainsi soumise à l'effet de l'âge, à l'origine des événements de non-disjonctions chromosomiques observés notamment dans le syndrome de Down. En revanche, la spermatogenèse comporte tout au long de la vie adulte une phase mitotique importante, de l'ordre de 23 divisions par an pour chaque spermatogonie, à l'origine d'une plus grande probabilité de variants *de novo* dans les gamètes mâles avec un effet de l'âge également plus important. On estime ainsi que chaque année supplémentaire par rapport à un âge de conception paternel de 20 ans augmente de 1 à 3 le nombre de variants *de novo* [235].

Les réarrangements de grande taille (CNV) survenant *de novo* sont estimés à une fréquence moindre, de l'ordre de 4.10^{-10} , soient environ 3 nouveaux CNV par génération et par génome [236]. Les délétions sont plus fréquentes que les insertions dans un rapport de 1:3, la fréquence de ces CNV est inversement proportionnelle à leur taille (0,01-0,02/par génération pour un CNV >100Kb), et à nouveau largement influencée par l'âge paternel.

Ces variants *de novo* peuvent également survenir en post-zygotique, et ainsi ne concerner que certains lignages cellulaires à l'origine de phénomènes en mosaïque. Ces situations complexifient la détection de ces événements *de novo*, nécessitant des approches par séquençage de grande profondeur pour être en capacité de détection de ces mosaïques sur prélèvement sanguin.

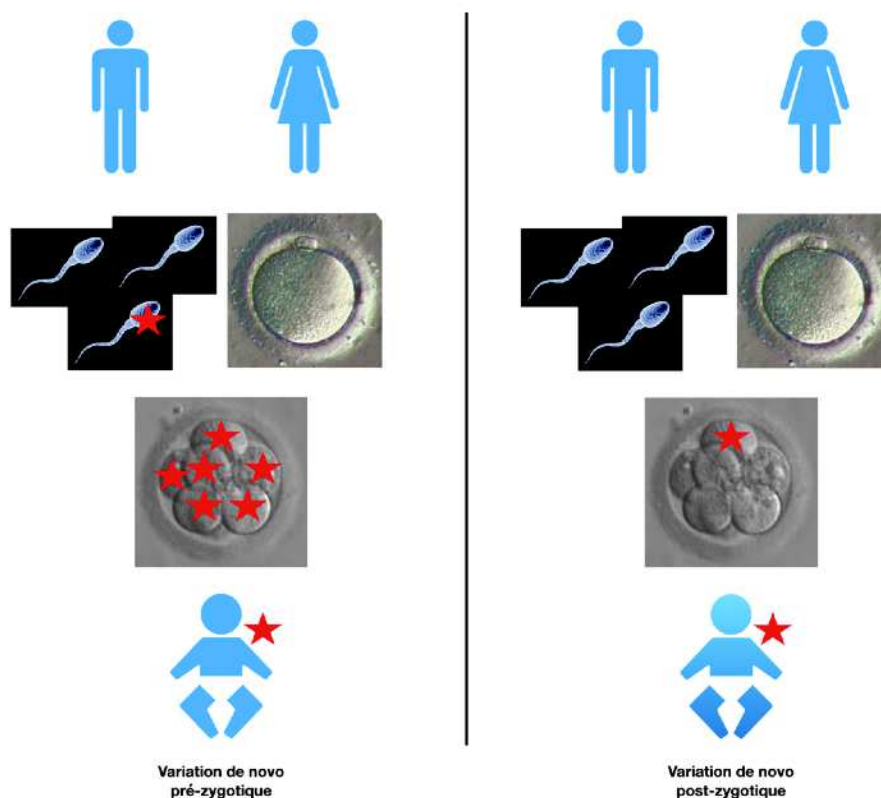


Figure 4 : Situations conduisant à la survenue de variations *de novo*

Ces altérations *de novo* ont révolutionné l'approche génétique, en proposant une origine génétique à des syndromes rares mais surtout en apportant un diagnostic moléculaire à certaines pathologies neuro-psychiatriques telles que la déficience intellectuelle, les troubles du spectre autistique ou la schizophrénie [232, 237, 238]. Ces variants *de novo* réconcilient également les théories évolutionnistes (*purifying selection*), puisqu'elles expliquent pourquoi l'incidence de pathologies génétiques graves, affectant de ce fait la capacité des individus atteints à se reproduire, demeurerait stable dans la population. L'approche séminale proposée par Vissers *et al* repose sur l'analyse de trios parents - enfant atteint, avec soustraction d'exomes à l'origine de l'identification de variants *de novo* possiblement causaux pour le phénotype observé, en l'occurrence ici des déficiences intellectuelles [237]. Il s'agit de ce fait d'un changement de paradigme, puisque ces approches en trios conduisent à identifier non pas des variations constitutionnelles héritées, mais des mutations constitutionnelles acquises aux stades pré ou post-zygotiques. La contribution de ces variations à diverses conditions pathologiques a de ce fait été largement explorées [239], en premier lieu pour les pathologies pédiatriques (Figure 5).

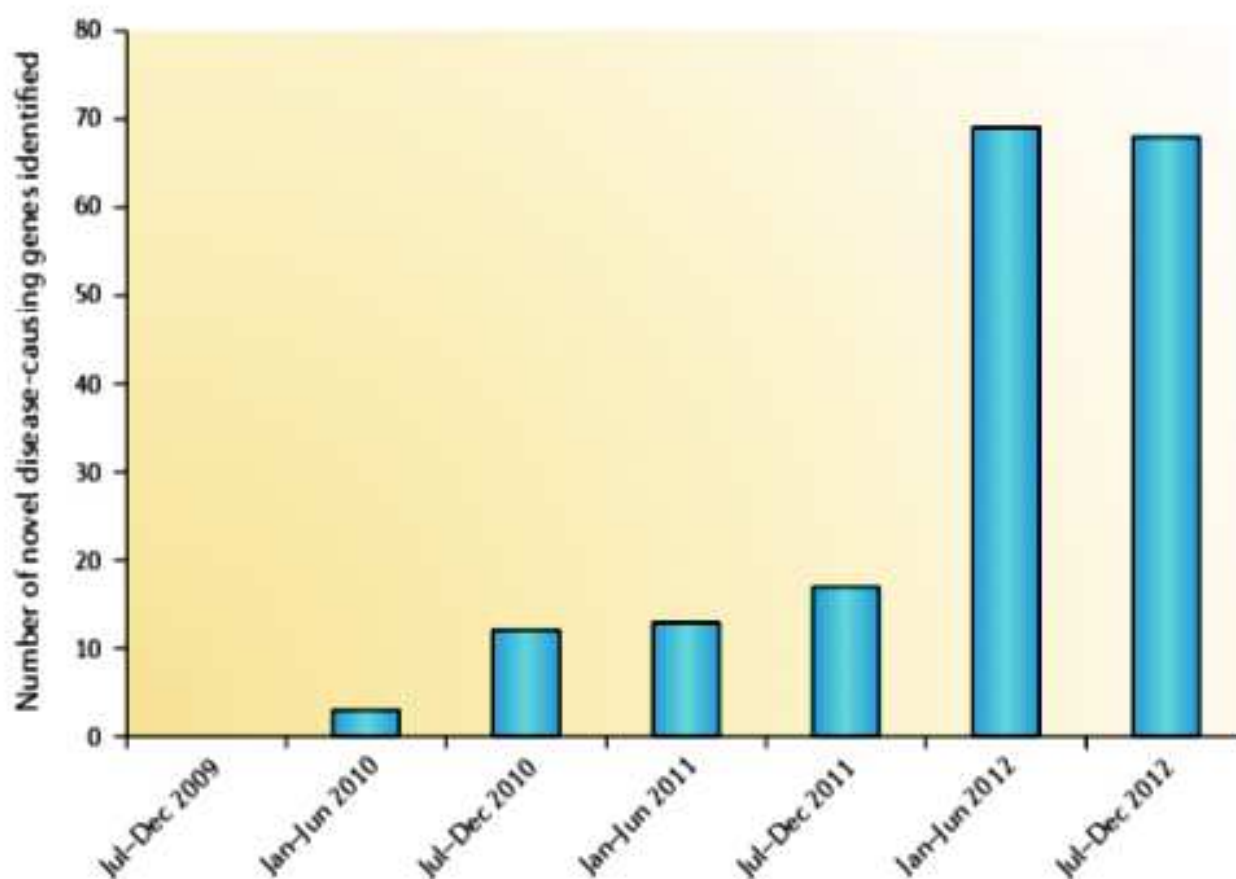


Figure 5 : identification des causes moléculaires de pathologies génétiques au cours du temps. Boycott *et al*, Nature Review in Genetics 2013 [239]

La présentation familiale dans ces situations est ainsi attendue comme sporadique, compte tenu de la sévérité des pathologies et de leur impact sur la capacité du sujet atteint à se reproduire et ainsi transmettre ce génotype acquis. Il s'agit de ce fait d'une seconde rupture de paradigme, les approches antérieures basées sur la récurrence et les études de liaison s'intéressant avant tout à des situations familiales lourdement affectées. Les concepts de **précocité** (*early-onset disease*) et de **phénotype extrême** (pathologie remarquable par une précocité et/ou une agressivité / sévérité inhabituelle) deviennent ainsi les portes d'entrée plutôt que le concept d'histoire familiale. L'implémentation du séquençage WGS ou WES dans le diagnostic moléculaire de ces situations conduit dans 60 à 75% des cas à l'identification d'une altération causale [240, 241]. Il s'agit pour l'essentiel d'altérations répondant à un mécanisme dominant [235]. L'autre élément remarquable est la faible récurrence de ces altérations parmi les sujets atteints, mais plutôt un ensemble de variants privés définissant une extraordinaire hétérogénéité génétique. Cette hétérogénéité peut néanmoins être clivée entre des pathologies relativement communes telles que les troubles autistiques, la schizophrénie ou la déficience intellectuelle où le nombre de gènes impliqués (*mutational target*) est important, et des pathologies génétiques plus rares où le panel de gènes incriminés est réduit à un ou quelques gènes.

Le champ pathologique le plus largement exploré et expliqué par ces variations *de novo* est constitué des pathologies neurologiques [238] et psychiatriques [242, 243]. La contribution de ces variants *de novo* dans la survenue de cancers est également connue, avec notamment 7% des variations constitutionnelles de *TP53* répondant à une situation acquise [115]. Une variation *de novo* correspondant à une variation d'épissage tronquante (c.423+1G>A) de *SDHB* a également été rapportée dans une situation de paragangliome pédiatrique [244]. Néanmoins, les cancers de phénotype extrême par leur précocité ou leur sévérité n'ont pas fait l'objet d'études extensives par cette approche en trios d'exomes. La difficulté majeure à nouveau réside dans l'interprétation et la démonstration de la causalité de ces variations *de novo* observées, compte tenu notamment de la probabilité forte à l'état physiologique de ces variations *de novo* ($1,18 \cdot 10^{-8}$ par position nucléotidique et par génome haploïde), avec un risque non négligeable pour chacune d'elle d'affecter un gène important, la séquence codante et ainsi la production d'une protéine fonctionnelle. Certaines variations *de novo* ont apporté une preuve irréfutable de leur pathogénicité, de part un phénotype homogène, un monogénisme, un gène affecté en relation avec la pathologie et surtout une récurrence au niveau d'individus atteints. Ainsi, le syndrome de Kabuki (MIM:147920) associe un retard mental congénital à un syndrome dysmorphique permettant cette homogénéisation phénotypique. L'identification de variations *de novo* pertes de fonction dans le gène *KMT2D* chez 9 sujets atteints parmi 10 analysés constituent des éléments irréfutables [245]. A l'inverse, les troubles du spectre autistique constituent une entité nosologique par définition hétérogène, et malgré de très larges séries de patients analysés, l'identification de variants *de novo* chez ces sujets se heurte à la question de leur imputabilité dans le phénotype. La probabilité de variation *de novo* pour un gène donné, une position donnée, n'est pas distribuée de manière homogène sur l'ensemble du génome ni corrélée à la taille des gènes, mais est

fonction d'un contexte génétique évoqué précédemment (richesse en îlots CpG, séquences riches en GC...). Cette probabilité pour un gène donné a été modélisée par un score d'intolérance (*Residual Variation Intolerance Score*) déterminé à partir des données exomiques de l'ESP6500 renseignant pour chaque gène une proportion variable gène-dépendant de variations rares ou polymorphiques présumées neutres [246]. L'hypothèse, basée sur des théories de conservation sélective, repose sur un rôle plus critique des gènes moins tolérants à ces variations *de novo*, et de ce fait à un argument de pathogénicité lorsqu'une variation affecte ces gènes. Un score de contrainte Z correspondant à la déviation entre les variations observées et celles statistiquement attendues, est donc proposé pour chaque gène [247] et disponible dans ExaC notamment. Les gènes les plus contraints correspondent ainsi sans surprise aux gènes impliqués dans les pathologies génétiques de transmission dominante, les gènes avec contraintes modérées étant plutôt impliqués dans les pathologies de transmission récessive. De manière similaire, un second score de contrainte appelé *pLI score* détermine pour chaque gène la tolérance aux variations pertes de fonction (LOF) [248]. Ces deux scores constituent ainsi, au-delà de la caractérisation et des critères de l'ACMG-AMP pour l'interprétation de la pathogénicité, des outils pertinents pour l'interprétation de variants *de novo* affectant des gènes fonctionnellement éloignés des mécanismes physiopathologiques connus pour la pathologie investiguée. Samocha *et al* appliquent ainsi ce modèle d'analyse aux variants LOF *de novo* identifiés chez des sujets atteints de troubles du spectre autistique et parviennent ainsi à discriminer parmi les gènes affectés par des variants *de novo* récurrents (définis par ≥ 2) les gènes causaux [247]. La base de données gnomAD (Genome Aggregation Database) (<https://gnomad.broadinstitute.org>) agrège les données de séquençage exomique de 125.748 individus issus de la base ExaC et les données de séquençage génomique de 15.708 individus, aboutissant à une base de données publiques de plus de 140.000 individus. A partir de cette base de données ainsi enrichie, une meilleure prédiction de la probabilité de pathogénicité d'un variant perte de fonction (pLOF) est obtenue par le ratio *o/e* (*observed / expected*), exprimé sous le terme LOEUF (*Loss-of-function Observed Expected Upper bound Fraction*). Cette métrique discrimine ainsi parfaitement les gènes avec haplo-insuffisance connue, les gènes impliqués dans des pathologies à l'état récessif et les gènes connus pour une tolérance maximale aux pertes de fonction (gènes olfactifs par exemple). La variable LOEUF est également parfaitement corrélée à l'intolérance bien documentée des gènes essentiels, définis soit sur la base de modèles murins de *knockout* avec létalité embryonnaire, soit approche d'extinction transcriptionnelle dans des modèles cellulaires [249].

L'interprétation des variants peut également s'appuyer sur des tests d'agrégation (*burden test*) dont le concept repose sur une sommation des variants observés pour un gène donné dans une population d'individus exposés au phénotype, pour ensuite comparer ce même variable chez des cas témoins. Ces scores permettent ainsi de prioriser par gènes et non plus par variants, et ainsi d'augmenter la puissance des tests dans le calcul des *odds ratios*. Différents test sont disponibles (KBAC, SKAT-O, VT ou VAAST) permettant de pondérer notamment par les fréquences observées en population contrôle [250]. Si l'on considère un gène X, large et donc

potentiellement cible d'un nombre important de variations ponctuelles ou de CNV, affecté chez un individu présentant un phénotype d'intérêt par un variant A rare, *de novo*, faux-sens et prédit délétère par des algorithmes bio informatiques, ce variant présenterait indéniablement des arguments de pathogénicité. Cependant, si nous observons dans la population contrôle pour ce même gène X, chez 50% des individus, d'autres variants prédits délétères, alors le *burden test* nuancera cette probabilité de pathogénicité.

Ainsi, l'interprétation d'un variant identifié dans un contexte phénotypique particulier et une population définie sur un plan ethnique, doit faire l'objet d'une démarche scientifique rigoureuse exploitant l'ensemble des ressources disponibles (Figure 6) : conséquences du variant au niveau de l'ARN ou de la protéine, données de conservation nucléotidique et protéique, prédictions bioinformatiques (SIFT, PolyPhen-2, MutationTaster...), implication du gène altéré dans une voie biologique cohérente avec le phénotype observé, analyses fonctionnelles (*in vitro*, modèles murins...), bases de données clinico-biologiques (Clinvar, HGMD, UniProt, Cosmic, TCGA...), fréquence du variant en population contrôle et contraintes du gène (Z-score, Pli-score, LOEUF) appréciables dans les bases de données (ExaC, gnomAD, ESP, 1000 Genome), données de récurrence du variant dans la population étudiée et recherche d'enrichissement par *burden tests* [251].

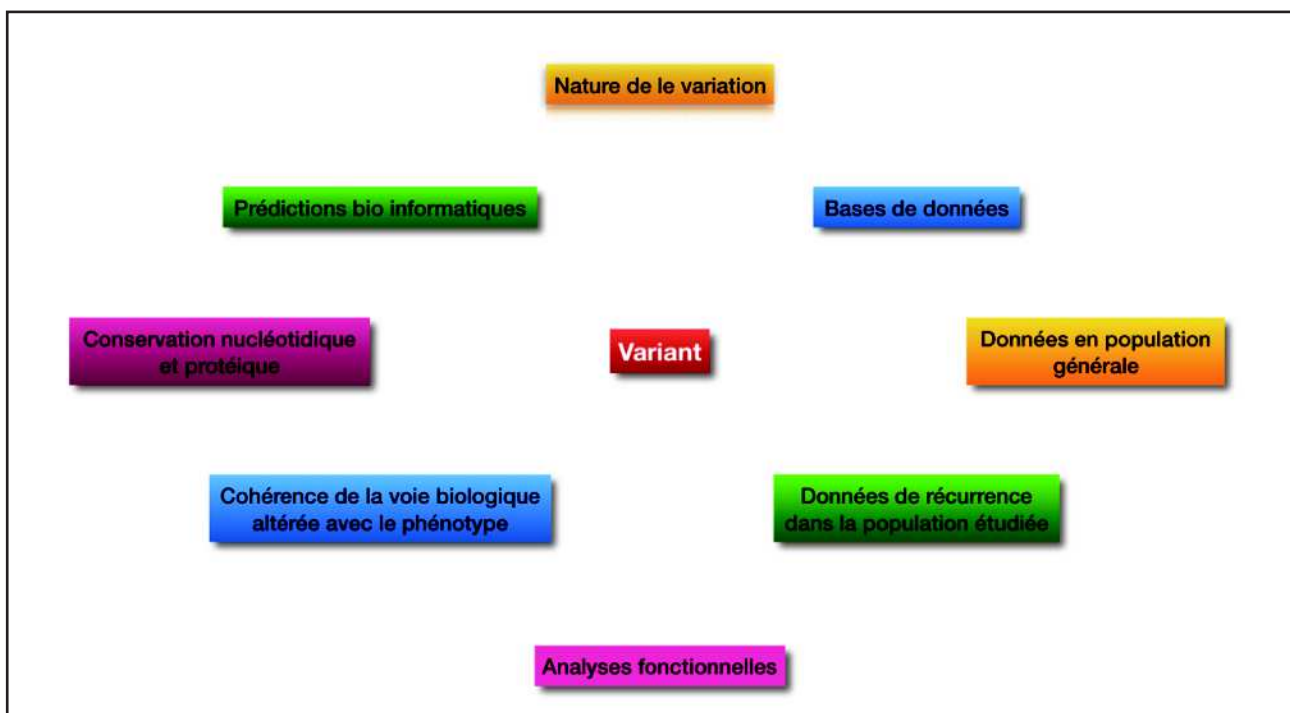


Figure 6 : Interprétation d'un variant

Objectifs de notre travail de Thèse

Notre approche se focalise donc sur les **cancers de présentation phénotypique extrême**, définie sur la base d'une extrême précocité, d'une sévérité particulière et/ou d'une multiplicité des primitifs, afin d'identifier des altérations constitutionnelles causales et ainsi élargir la connaissance des bases moléculaires de la prédisposition au cancer. Les outils employés sont d'une part, une large collection biologique d'ADN de sujets cas index analysés antérieurement et avec résultats négatifs pour les gènes canoniques de la prédisposition, et d'autre part les techniques de séquençage de nouvelle génération ainsi que les algorithmes de filtration bioinformatique développés au sein de l'unité Inserm 1245. Nos travaux se sont déclinés en trois parties (Figure 7) : nous avons dans un premier temps utilisé une stratégie d'**exomes soustractifs ou comparatifs**, puis dans une seconde approche par **panel de gènes élargis** aux voies de la cancérogenèse, et enfin dans un troisième temps nous nous sommes focalisés sur la problématique des **variants de novo en mosaïques**.

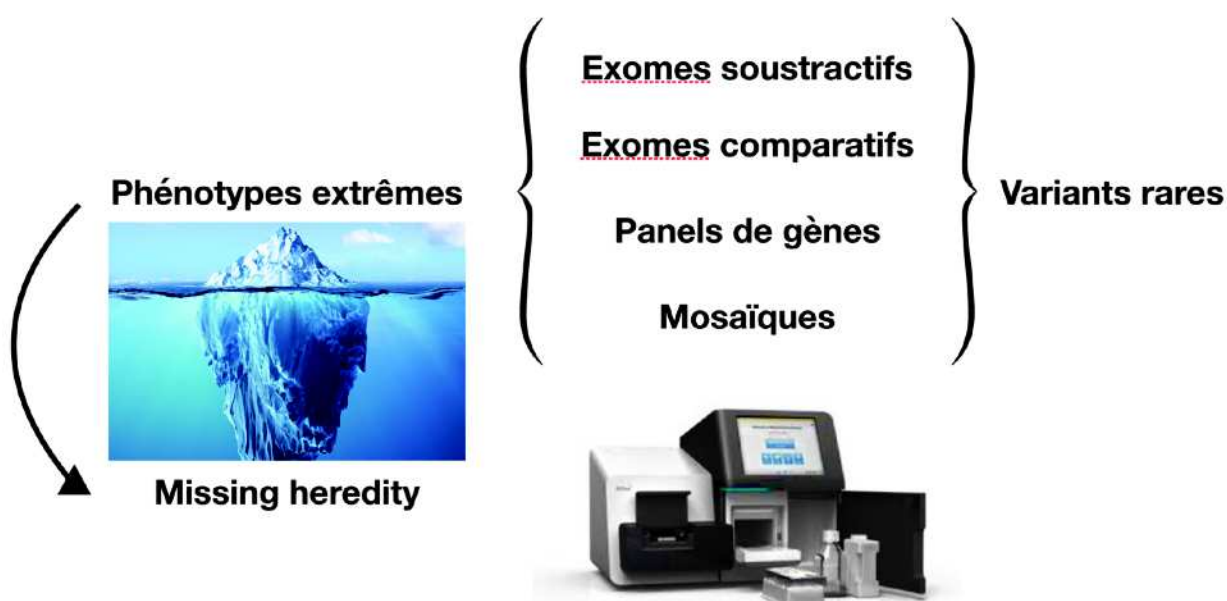


Figure 7 : Stratégies déployées et inférences de la contribution de variants rares identifiés

Partie I : Exomes soustractifs et comparatifs dans les cancers du sein et de l’ovaire de phénotype extrême.

Contexte et rationnel

Comme développé précédemment, une large fraction de situations cliniques et familiales suspecte de déterminisme génétique à l’origine de la survenue de cancers du sein ou de l’ovaire, reste inexploquée par les analyses génétiques de *BRCA1*, *BRCA2*, *TP53*, *PALB2*, *RAD51C* et *RAD51D*, voire de manière plus extensive par des panels de gènes explorant plus largement la voie de la réparation de l’ADN. L’extrême précocité des cancers du sein définie comme survenant avant 31 ans, ou des cancers de l’ovaire défini comme survenant avant 41 ans, apparaissent comme des marqueurs forts d’un déterminisme génétique. La stratégie des **exomes soustractifs** sur des trios sélectionnés pour un cas index de présentation extrême et des parents (et plus largement des apparentés aux 1er et 2ème degré) indemnes, apparaît comme une stratégie pertinente pour identifier des variants *de novo* à l’origine du phénotype observé. Il s’agit ici d’appliquer une stratégie performante dans le champ des troubles neuro-psychiatriques, à la manière de Vissers *et al* [237]. Dans une autre approche, nous avons sélectionné une famille remarquable par l’agressivité et la précocité des cancers observés, avec des cas index distants de plusieurs degré de parenté, permettant une approche de type **exomes comparatifs**.

Exomes soustractifs dans un trio parent / enfant affecté par un cancer de l’ovaire précoce.

Contexte clinique

Notre premier trio analysé s’intéressait à une présentation dramatique de cancer de l’ovaire chez une jeune femme de 21 ans. Sur le plan anatomo-pathologique, la tumeur ovarienne correspondait à un adénocarcinome micro-papillaire de bas grade (grade 1 selon Silverberg ou bas grade selon la classification du M.D. Anderson), diagnostiquée en 2010 à un stade avancé (stade IIIc selon la classification FIGO) avec atteinte ovarienne bilatérale et carcinose péritonéale étendue infiltrant le colon sigmoïde, le caecum, ainsi qu’une atteinte ganglionnaire pelvienne et lombo-aortique étendue. Le premier temps thérapeutique a comporté une chirurgie de *debulking* majeure avec hystérectomie et annexectomie bilatérale, pelvectomie postérieure avec résection colorectale, hémicolectomie droite, omentectomie infra-gastrique, résections péritonéales multiples notamment au niveau de l’hémi-diaphragme droit, curages ganglionnaires extensifs. La patiente a bénéficié d’une chimiothérapie adjuvante par Paclitaxel - Carboplatine 6 cures puis d’une tentative de rétablissement de la continuité. Il est alors constaté la récurrence de la maladie sur l’orifice stomial, et l’évolution va être dès lors rythmée par des occlusions digestives à répétition sur la carcinose péritonéale, difficilement jugulées par plusieurs lignes de chimiothérapie à base de sels de platine, de taxanes, de Gemcitabine, de Topotécan,

d'anthracyclines. La patiente répondra transitoirement (12 mois) à une hormonothérapie par Tamoxifene, cohérente avec le bas grade observé sur la tumeur initiale. La patiente décèdera des complications d'un sepsis sévère, six ans après le diagnostic initial.

Sur le plan de l'histoire familiale, la patiente est issue de deux parents indemnes, respectivement âgés de 36 ans et 31 ans pour le père et la mère, lors de la naissance du cas index. Un frère plus âgé de 30 ans est également indemne. Les deux branches parentales ne présentaient aucun historique de cancers, avec cependant une branche paternelle peu documentée. Les analyses *BRCA1* et *BRCA2* sont négatives, ainsi que l'analyse du gène *TP53*.

Compte tenu de cette présentation atypique par l'âge d'apparition inhabituel pour un cancer de l'ovaire et son agressivité particulière, la patiente et ses parents ont bénéficié d'une analyse en exomes afin d'effectuer une analyse de type soustractive.

Matériel et Méthode

L'ADN constitutionnel a été extrait de prélèvements sur sang périphérique pour les deux parents. Pour le cas index, dans une perspective également d'identification d'altération en mosaïque, l'ADN constitutionnel a été extrait d'un prélèvement tissulaire ovarien sain effectué lors de l'intervention initiale.

La filiation du cas index a été vérifiée par étude de 4 micro-satellites autosomaux (D1S439, D9S1784, D14S986, D19S913) et 3 micro-satellites portés sur le chromosome X (DXS 16, DXS 8022 et DXS 8028).

La préparation des librairies d'ADN a été effectuée selon le protocole *SureSelect^{XT} Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library* (Agilent Technologies) reposant sur (Figure 8) :

- (i) : fragmentation par sonication de l'ADN avec 3µg d'*input* de qualité vérifiée par fluorométrie (*TwinckleTM LB970, Berthold, France*), en molécules de 200 paires de bases. La sonication a été effectuée à l'aide de l'appareil Covaris S220 (*Woburn, MA 01801 US*).
- (ii) : réparation des extrémités cohésives de l'ADN fragmenté par enzyme de *Klenow* avec obtention d'extrémités franches grâce aux activités exonucléases 3'→5' et polymérase 5'→3'. Ajout d'une molécule d'ATP aux extrémités 3'.
- (iii) : ligation des adaptateurs comportant les séquences complémentaires aux *primers* permettant l'amplification par PCR, ainsi que les séquences permettant la fixation à la *flowcell* des produits amplifiés.
- (iv) : Amplification par PCR par 4 à 6 cycles avec dénaturation 98°C 30 secondes, hybridation 65°C 30 secondes et extension 72°C 1 minutes.

Chaque étape est suivie d'une purification sur billes magnétiques (*AMPure XP beads*). Un contrôle de la qualité de la sonication est effectué sur TapeStation 4200 (*Agilent, Santa Clara, CA 95051, USA*), afin de vérifier la taille et la quantité des fragments issus de la sonication, avant et après amplification.

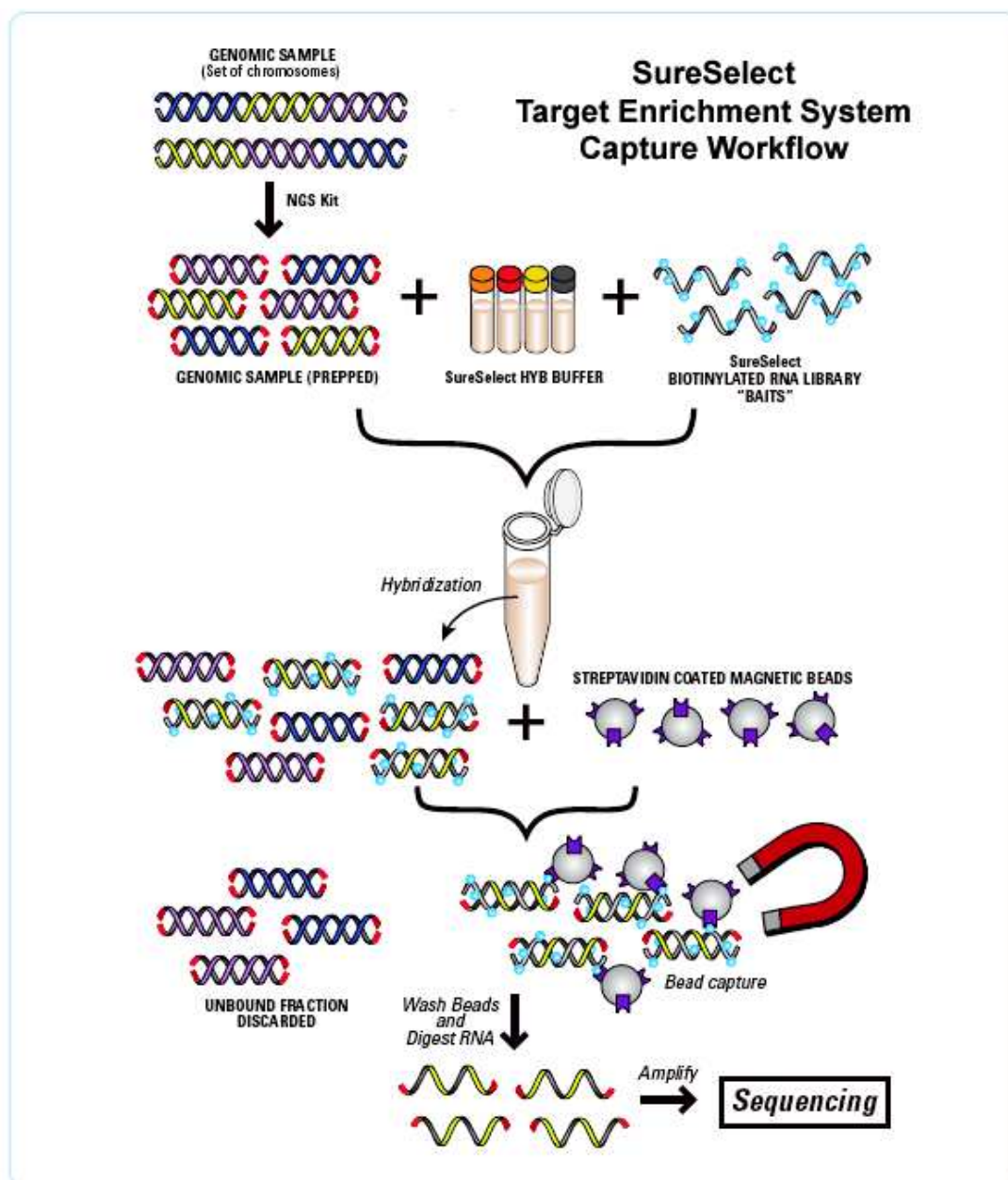


Figure 8 : Protocole SureSelect Agilent et préparation des bibliothèques d'ADN

La capture des exomes a été effectuée par hybridation, selon la méthode développée par Tewhey *et al* [252] et le kit *SureSelect Human All Exons 50 Mb kit* (24000 genes, Agilent Technologies). La capture repose sur l'hybridation de sondes d'ARN biotinylées et complémentaires des régions d'ADN d'intérêt, en l'occurrence de l'ensemble des séquences exoniques. Les ADN préparés sont ramenés à une concentration de 221 ng/μL dans un volume de 3,4 μL. L'hybridation est effectuée en thermocycleur avec le kit de capture pendant 24h à 65°C. Le produit hybridé est capturé au décours par des billes de streptavidine. Les index de multiplexage associés aux *primers* de séquence sont insérés avant nouvelle amplification par PCR 10 à 16 cycles. Après nouvelle purification sur billes magnétiques, les tailles et concentrations des fragments hybridés et amplifiés sont contrôlées sur TapeStation. Une mesure

précise de la concentration de chaque index est effectuée par qPCR avant la constitution d'un pool équimolaire.

Le séquençage des exomes a été effectué sur le séquenceur Illumina *Genome Analyzer II* (GAIIx) localisée au sein de l'unité Inserm 1245. La première étape consiste à déposer sur la *flowcell* les ADN capturés, qui s'hybrident avec les amorces oligonucléotidiques fixées sur la lame et complémentaires des régions d'intérêt. Une amplification en cluster permet d'obtenir un séquençage de forte profondeur, par la répétition de cycles comportant : synthèse à partir de l'extrémité 3' d'une séquence de 75 paires de bases complémentaires du brin hybridé à partir de l'amorce, dénaturation du brin matrice et hybridation en pont du brin néosynthétisé avec l'amorce adjacente, synthèse du brin complémentaire à partir de l'amorce adjacente. Une nouvelle dénaturation conduit à deux molécules simples brins antiparallèles fixées chacune sur une amorce. Aux termes de ces cycles et après clivage des brins anti-sens, plusieurs milliers de copies sont générés à partir d'un simple brin d'ADN (Figure 9). Le séquençage est effectué en mode *paired-end*, signifiant que chaque fragment est séquencé deux fois, à partir de l'extrémité

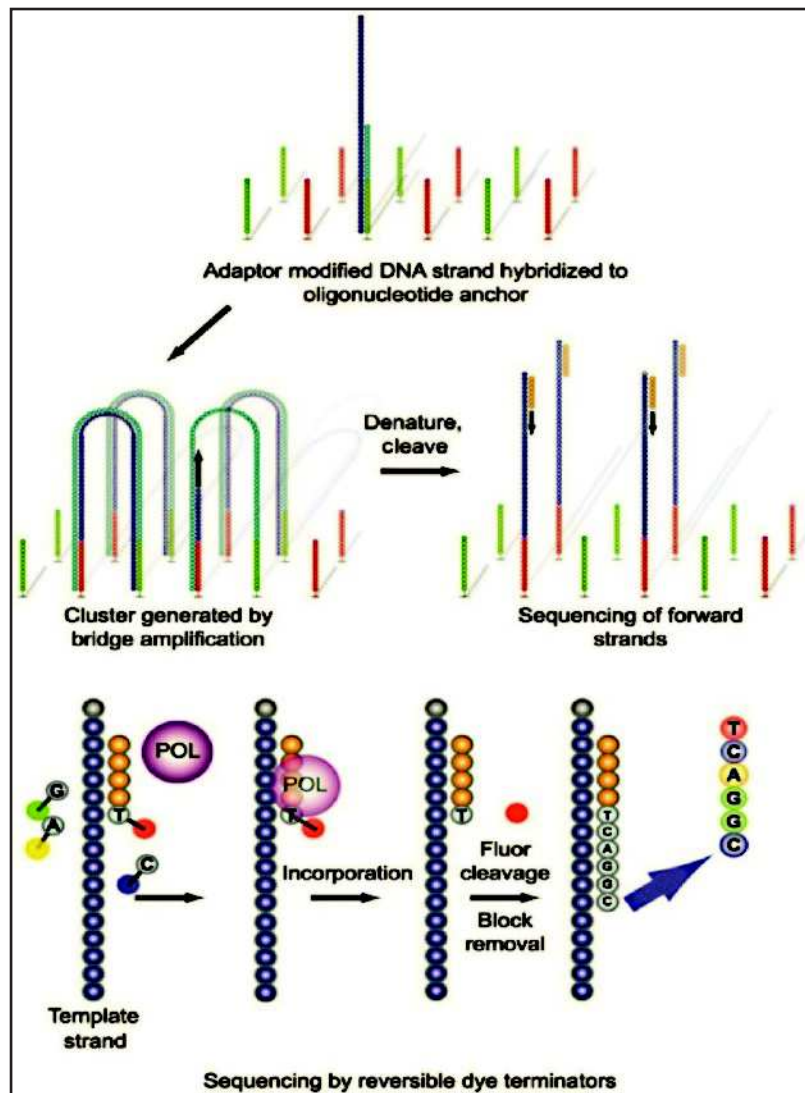


Figure 9 : Formation des clusters - technologie Illumina

3' du brin sens et anti-sens, par une technique de réversion des clusters. L'intérêt de ce séquençage *paired-end*, outre de multiplier par 2 la profondeur, permet une amélioration de l'alignement, la détection des *indels*, et l'élimination des duplicats de PCR.

Le pipeline bioinformatique comporte schématiquement six étapes, réparties en analyse primaire et secondaire. L'analyse primaire est effectuée par le logiciel RTA 1.10 d'Illumina basé sur un analyseur d'image permettant un premier contrôle qualité des données et l'opération de *base calling*, soit la détermination des séquences produites selon l'intensité lue des fluorochromes associés à chaque analogue nucléotidique inséré. Un score de qualité est associé à chaque base appelé score de Phred, correspondant à une probabilité d'identification erronée d'une base à cette étape. Le score est exprimé selon une base logarithmique, soit pour un score de 10 une probabilité d'erreur de 10%. Un score agrégé de qualité est fourni sous la forme du Q-score, un Q-score de 30 (Q30) correspondant à un risque d'erreur de 1/10000. Un fichier .bcl est généré au terme du *base calling*. Les quatre étapes suivantes correspondent à l'analyse secondaire, et ont été effectuées à l'aide d'outils distincts de la solution Illumina. Le démultiplexage correspond à la lecture des index afin de générer pour chaque patient un fichier FASTQ comportant les données de séquence et les scores de qualité associés. L'alignement correspond à la recherche du maximum de similitudes entre les séquences générées et un génome de référence. Un fichier BAM est ainsi généré au terme de l'alignement. Plusieurs outils sont disponibles pour gérer cette étape, nous avons utilisé ici le logiciel CASAVA d'Illumina version 1.8.2, et pour le génome de référence Human hg19/GRCh37. La détection des variations par rapport au génome de référence (*variant calling*) est également médiée par différents outils, ici également par CASAVA. Un fichier .vcf est généré à l'issue de cette détection, et annoté selon la nomenclature HGVS en intégrant les informations relatives à la position génomique, le gène impliqué, le type de variation, le statut hétéro- ou homozygote, la fréquence allélique de l'allèle alternatif, la profondeur de lecture et le score de qualité associé, la fréquence rapportée dans les données de population, le rs associé si décrit dans dbSNP, l'effet produit par les algorithmes de prédiction. Une double annotation par ANNOVAR et *Variant Effect Predictor* a été ici utilisée.

La filtration de ces variants est l'élément clé dans le séquençage de nouvelle génération, compte tenu de la quantité de données générées. Ainsi, pour un exome, le nombre de variations attendu est autour de 20.000. Nous avons utilisé dans cette stratégie d'exomes soustractifs une solution développée en interne appelée *Exome Variant Analyser* (EVA) développée par une bioinformaticienne (Sophie Coutant) de l'unité Inserm 1245 [253]. Cette interface de filtration permet de réduire par des étapes contrôlées le nombre de variants, en :

- (i) éliminant les variants connus non rares référencés dans les bases de données telles que dbSNP, 1000 Genomes Project, Exome Sequencing Project (ESP), Complete Genomics Public Data ou Integrative Genomics Public Data.
- (ii) filtrant les variants selon leurs localisations géniques (introniques / exoniques / 3' ou 5' UTR), les catégories fonctionnelles (faux-sens, non-sens, synonymes, *indels frameshift* ou non), les scores de qualité et la couverture / profondeur.

(iii) appliquant une stratégie de filtration particulière telle que la recherche de récurrence parmi des familles indépendantes, la recherche d'homozygotes / hétérozygotes / hétérozygotes composites au sein d'une même famille, la recherche de variants *de novo* dans une approche par trios parents / enfant atteint.

La stratégie appliquée ici est bien entendu celle des exomes soustractifs, en retranchant les variants communs entre le cas index et les deux parents pour ne conserver que les variants *de novo* chez le cas index, de type extrêmement rares ($<0,1\%$), exoniques, associés à une profondeur d'au moins 10x, un Q-score supérieur à 30.

La visualisation des séquences sur fichier BAM a été effectuée sur le logiciel Alamut (Interactive Biosoftware) pour l'interprétation finale et l'élimination des artéfacts.

Les variants identifiés et conservés ont été confirmés par une méthode Sanger utilisant le *BigDye® Terminator v3.1 Cycle sequencing Kit* (Life technologies) et le séquenceur *3100 xl Genetic Analyzer* (Life technologies).

Résultats

La confirmation de la filiation du cas index a été obtenue par étude des micro-satellites. Les exomes générés à partir des ADN des deux parents et du cas index ont permis d'obtenir en moyenne 8 Gigabases de données par exome, avec une profondeur moyenne de 69x et 88% des bases présentant une profondeur minimale de 10x, un alignement au génome de référence pour 98% des *reads*, un Q-score supérieur à 30 pour 89% des bases. Une moyenne de 17.395 variants exoniques a été obtenue par exome. Nous avons dans un premier temps filtré les variants sur des critères de qualité : couverture des *reads* supérieure à 10x et Q-score supérieur à 30. Les BAM correspondant aux gènes canoniques de la prédisposition héréditaire au cancer de l'ovaire (panel HBOC) ont été vérifiés manuellement. Les variants rares de $MAF < 0,1\%$ ont été seuls conservés, en excluant tous les variants de fréquence allélique $> 0,001$ rapportés dans les bases *1000 genome project* et *ESP*. Les variations présentes à l'époque dans la base de données interne à l'inserm U1245 (72 exomes issus d'individus atteints de maladie d'Alzheimer précoces) ont permis une filtration supplémentaire. Ces premiers niveaux de filtration ont permis de retenir à ce stade en moyenne 306 nouveaux variants par individus. La soustraction des variants parentaux aux variants du cas index a conduit à retenir 46 variants résiduels présumés *de novo*. Une filtration manuelle a alors été appliquée par visualisation des BAM, avec 37/46 variants présents à faible balance allélique chez les parents. Parmi les 9 variants résiduels, 8 apparaissaient artéfactuels à la lecture des BAM (balance allélique inférieure à 20/80) et non confirmés par séquençage Sanger. La variation c.1157A>G/p.Asn386Ser du gène *INHBA* (NM 002192.2; MIM #147290) a été ainsi identifiée comme *de novo* et confirmée par séquençage Sanger chez le cas index, et absente chez les parents (Figures 10 et 11). La confirmation chez le

cas index a de plus été effectuée sur le tissu ovarien non tumoral, le tissu ovarien tumoral et le sang périphérique, excluant une situation de mosaïque.

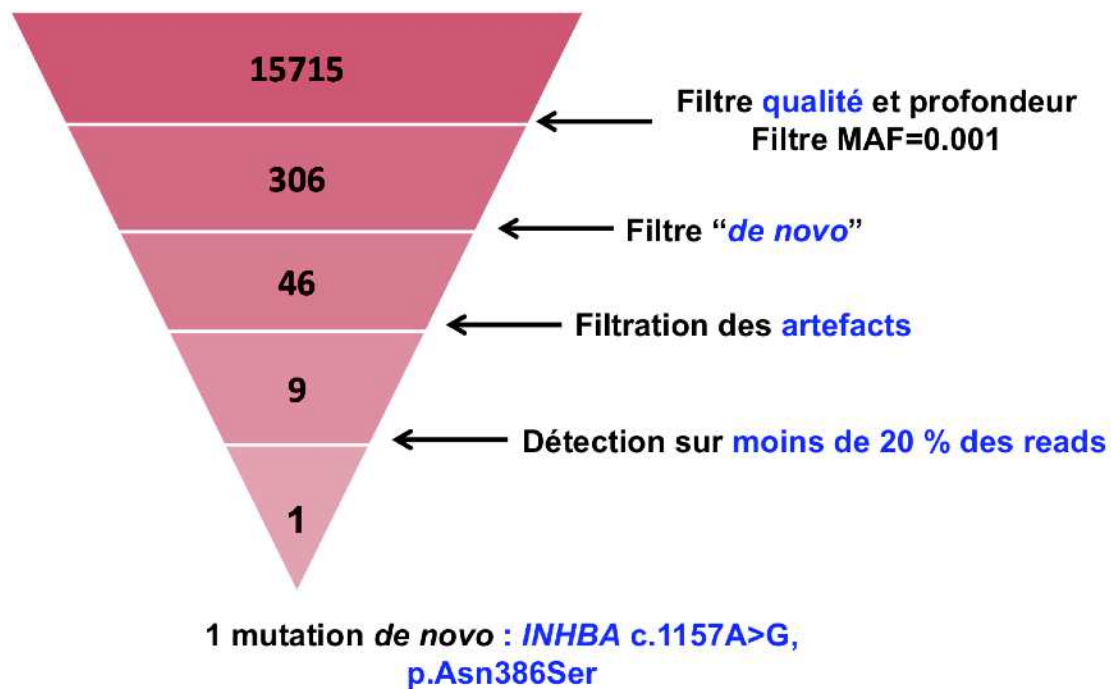


Figure 10 : Filtration des variants

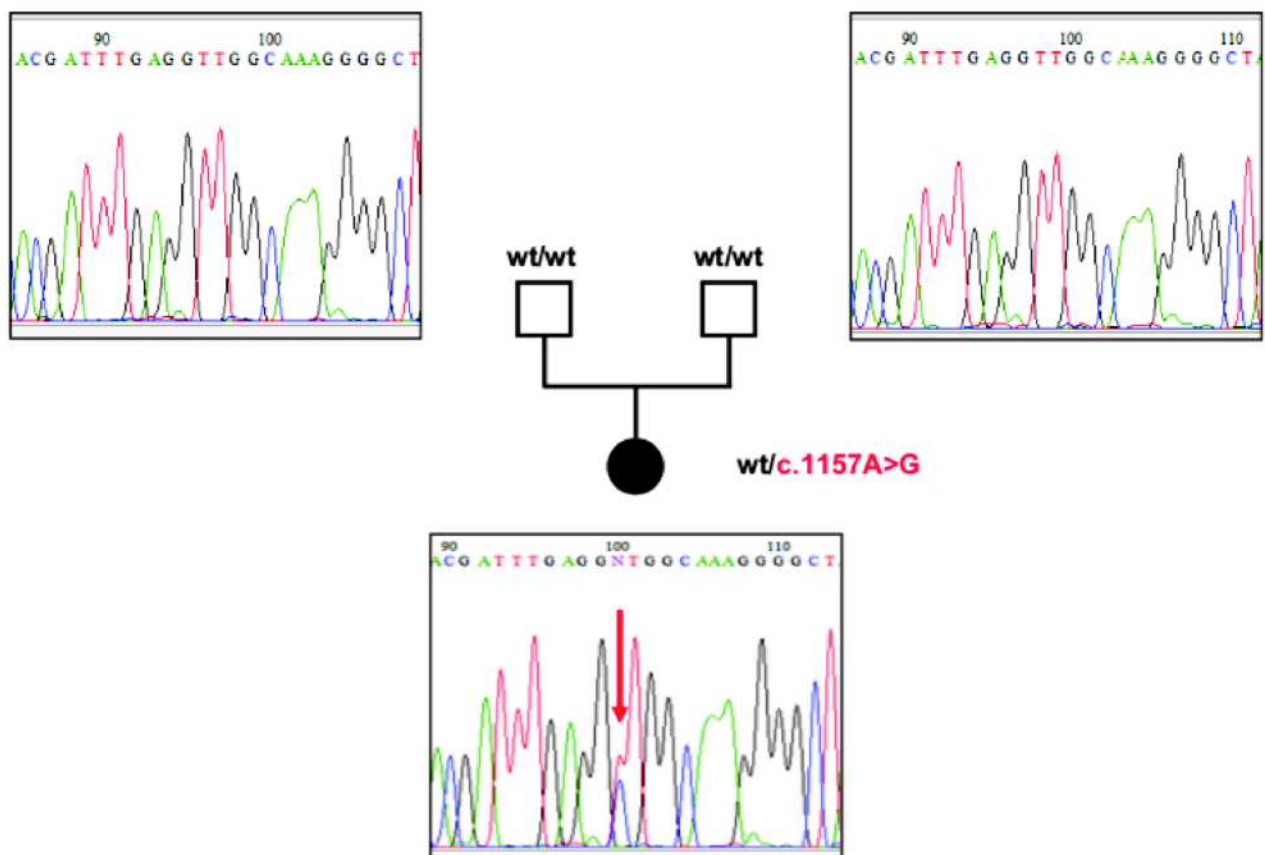


Figure 11 : Séquençage du gène *INHBA* par méthode Sanger du cas index et des parents

La variation c.1157A>G du gène *INHBA* est une variation faux-sens conduisant au remplacement d'un résidu asparagine par une sérine au niveau du codon 386. La distance de Grantham associée à cette substitution est 46, s'agissant de deux acides aminés de propriété physico-chimique semblable polaires non chargés. Cette variation n'était pas présente dans les bases *dbSNP*, *ESP6500*, *1000 Genome Project* et *ExaC*. Depuis 2017 est néanmoins apparu un référencement dans *dbSNP* sous le numéro rs1361491625 associé à une fréquence allélique de 0,00001. Le *Z score* pour les variations faux-sens pour *INHBA* est 2,72 et le *pLI score* 0,56 signifiant un degré modéré de contrainte sur ce gène. Les données gnomAD sont similaires avec un rapport *o/e* de 0,56 pour les variations faux-sens. Les algorithmes de prédiction indiquent pour cette variation des prédictions discordantes, avec pour *AlignGVGD* une classe C0 (GV 107,72 - GD 0,00), pour *SIFT* une prédiction de tolérabilité (score 0,31), pour *MutationTaster* en revanche une prédiction de pathogénicité (*Disease causing*) et pour *PolyPhen-2* une prédiction de bénignité (score 0,020). Ce variant ne présente pas d'effet prédit sur l'épissage, compte tenu de la position centrale de cette variation dans l'exon 3, ce dernier constituant par ailleurs le dernier exon de *INHBA*. Cette position est moyennement conservée, jusqu'au poulet, mais correspond au domaine fonctionnel carboxy-terminal de la sous-unité β_A de l'inhibine. Cette sous-unité β_A de l'inhibine interagit sous forme d'homodimères pour constituer l'Activine A, ou sous forme d'hétérodimères avec la sous-unité α de l'inhibine (codée par *INHBA*) pour constituer l'Inhibine A.

Ces gènes appartiennent à la superfamille du gène TGF- β , et sont impliqués dans la régulation ovarienne. L'inhibine et l'activine sont sécrétées par les cellules de la granulosa et participent à la régulation endocrine (rétro-contrôlant la FSH) et paracrine de l'épithélium ovarien et notamment de la folliculogénèse [254]. Ces deux protéines sont produites sous forme de précurseurs, avec un pro-domaine clivé après assemblage des homo ou hétérodimères. L'Activine se fixe sur les récepteurs sérine/thréonine kinase type II A ou B (ActRIIA / ActRIIB). Cette fixation initie le recrutement du récepteur Activine type I (ALK4 ou ALK7) et initie la cascade de transduction médiée par la voie Smad 2/3 aboutissant par un effet transcriptionnel à l'activation de voies de la prolifération [255]. L'Inhibine agit comme compétiteur sur les récepteurs ActRIIA/B en présence du co-récepteur betaglycan, régulant ainsi négativement la voie. La Follistatine participe également à cette régulation négative en liant l'Activine. L'implication dans la cancérogenèse ovarienne est fortement suspectée par l'observation de tumeurs dérivées des cordons sexuels chez les souris *knock-down* pour la sous-unité α [256], la perte d'expression de la sous-unité α et à l'inverse l'hyper-expression de la sous-unité β_A dans des lignées cellulaires de tumeurs ovariennes [257], conduisant à suspecter que l'Activine A puisse agir comme oncogène dans la carcinogénèse ovarienne.

Malgré sa nature faux sens et les discordances des algorithmes de prédiction, l'identification de cette variation faux-sens *de novo*, dans un gène encodant une protéine impliquée dans la régulation ovarienne, chez une jeune femme atteinte de manière exceptionnellement précoce par un adénocarcinome ovarien, nous a conduit à envisager que le gène *INHBA* était un sérieux gène candidat à la prédisposition génétique au cancer de l'ovaire. En

outre a été observé la perte d'hétérozygotie de l'allèle sauvage de *INHBA* au niveau de la tumeur, soulignant le rôle possiblement causal de ce gène dans la cancérogenèse survenue chez cette patiente.

Néanmoins compte tenu des difficultés d'interprétation et d'imputabilité des variants *de novo* développées précédemment, nous avons mené d'une part une recherche de récurrence au sein de cohortes de patientes atteintes de cancers de l'ovaire précoce, et d'autre part réalisé des approches fonctionnelles en collaboration avec une équipe australienne spécialiste de ce gène (Walton K. et Craig H. du Prince Henry's Institute of Medical Research à Clayton, Australia). Compte tenu du mécanisme d'homo ou hétérodimérisation des sous-unités α et βA , la recherche de variations récurrentes a concerné l'ensemble des séquences codantes tant du gène *INHBA* que celles du gène *INHA* (NM 002191.3). La cohorte a été composée de 43 patientes atteintes de tumeurs de l'ovaire survenant avant 40 ans et négatives pour les analyses *BRCA1* et *BRCA2*. Parmi ces 43 patientes, 14 provenaient de collections françaises (Centre François Baclesse - Institut Gustave Roussy - Institut Curie) et 29 provenaient d'une collection américaine (M.C. King, Department of Medicine and Genome Sciences, University of Washington School of Medicine, Seattle, Washington). Les analyses ont été effectuées par méthode Sanger avec séquençage complet des deux gènes compte tenu de leurs tailles. Aucune nouvelle variation prédite délétère n'a été retrouvée concernant le gène *INHBA*. En revanche, nous avons identifié une variation hétérozygote de *INHA* (MIM*147380) en position c.179G>T/p.Arg60Leu, chez une jeune femme de 29 ans ayant présenté une tumeur borderline de type cystadénome papillaire. Le caractère hérité ou *de novo* n'a pu être établi, ne disposant pas de l'ADN parental. La variation a en revanche été retrouvée au niveau de la tumeur. Cette variation faux-sens présente une distance de Grantham de 102 puisque modifiant un acide aminé basique en acide aminé hydrophobe, n'est pas rapportée dans les bases de données et est prédite délétère (*AlignGVGD C35, SIFT : deleterious, Mutation Taster : Disease causing, PolyPhen-2 : probably damaging*). Le résidu Arginine 60 est hautement conservé et contribue à une séquence consensus (Arg - X - X - Arg) permettant le clivage par une furine en position N-terminale de la sous-unité α . Nous avons amplifié la cohorte par 62 nouveaux cas de tumeurs ovariennes également précoces (<40 ans) afin d'identifier éventuellement de nouvelles variations prédites délétères ou un enrichissement éventuellement en variants rares pour ces deux gènes. Une variation faux-sens de *INHBA*, non rapportée dans les bases de données, a été détectée dans cette extension de cohorte, c.839G>A, p.Gly280Glu, chez une jeune femme affectée par un adénocarcinome séro-papillaire ovarien à 27 ans. En outre, 3 variations faux-sens de *INHA* (c.769G>A/p.Ala257Thr ; c.127C>G/p.Pro43Ala et c.487G>A/p.Val163Met) ont été retrouvées, correspondant à des variants rares (<1%) répertoriés dans les bases de données. De même, nous avons identifié une variation faux-sens de *INHBA* (c.686G>A/p.Arg229Gln) répertoriée avec une MAF <1%. L'ensemble des variants non-synonymes identifiés est résumé dans le tableau 7.

Afin d'apporter un poids statistique à ces variants identifiées, nous avons testé l'hypothèse d'un éventuel enrichissement en variants rares des gènes *INHA* et *INHBA* par rapport

Gene	rs	Nucleotide Variation	Predicted protein Variation	EVS6500 MAF (EA/AA/All)	1000 genomes MAF	Cases MAF
<i>INHA</i>	-	c.179G>T	p.Arg60Leu	Unreported	Unreported	0.006
<i>INHA</i>	rs12720062	c.769G>A	p.Ala257Thr	0.025/0.0045/0.018	0.01	0.053
<i>INHA</i>	rs12720060	c.127C>G	p.Pro43Ala	0.0006/0.0/0.0004	Unreported	0.006
<i>INHA</i>	rs149916845	c.487G>A	p.Val163Met	0.0019/0.00023/0.0013	0.0014	0.006
<i>INHBA</i>	-	c.839G>A	p.Gly280Glu	Unreported	Unreported	0.006
<i>INHBA</i>	rs138819536	c.686G>A	p.Arg229Gln	0.0035/0.0009/0.0026	0.0009	0.006

Tableau 7 : Variants rares non-synonymes identifiés dans la cohorte de cancers des ovaires précoces

à une population contrôle. Cette population contrôle a été constituée par extraction des données de séquençage des gènes *INHA* et *INHBA* issues de deux sources différentes, afin de réduire les possibles biais de séquençage et de population. La première cohorte est issue de la population *1000 Genomes (Public Database)* en sélectionnant des populations d'origines géographiques / ethniques simulant la diversité française mais reflétant l'origine caucasienne dominante de notre cohorte de réplication : *CEU* (résidents de l'Utah originaire majoritairement d'Europe Occidentale), *TSI* (originaires d'Italie), *FIN* (originaires d'Europe scandinave), *GBR* (originaires du Royaume-Uni) et *IBS* (originaires de la péninsule ibérique). Ce premier jeu de données correspond ainsi à 154 individus non apparentés, et permet l'identification de 5 variants faux-sens avec fréquence allélique inférieure à 1%. La seconde cohorte est extraite de *ESP6500*, avec sélection de 4300 individus hommes et femmes d'origine européenne. Les variants faux-sens d'*INHA* et *INHBA* dans ce second jeu de données s'élèvent à 30 variants, dont la majorité à l'exception d'un variant (rs12720062, MAF 2,5%) sont rares (MAF<1%).

Nous avons donc comparé les 7 variants identifiés dans le trio et la cohorte de réplication (n=76) avec les 5 variants identifiés dans la cohorte *1000 Genome* (n=154) et les 30 variants identifiés dans la cohorte *ESP6500* (n=4300). Afin de tester l'hypothèse d'un enrichissement en variants rares d'*INHA* et *INHBA* chez des patients affectés par des cancers de l'ovaire précoce (<40 ans), nous avons effectué différents tests en l'absence de consensus sur la méthodologie statistique pour ces variants rares. Les tests dits *burden tests* (**CAST** pour *cohort allelic sum test*, [258] et **CMC** pour *Combined Multivariate and Collapsing method*, [259]) agrègent et collapsent l'ensemble des variants rares en un variant unique fictif et testent l'association de ce variant virtuel avec le phénotype (patient ou contrôle). Les deux tests donnent de manière concordante des *p-value* proches de la significativité (p-valeurs calculées à l'aide du logiciel *AssotesteR R package*). Les tests C-ALPHA [260] et SKAT [261] sont utilisés pour leur caractère bidirectionnel permettant de prendre en compte l'éventuel effet protecteur ou neutre de ces variants. De nouveau, les tests apparaissent en faveur d'un enrichissement significatif en variants rares dans les gènes *INHA* et *INHBA* chez les patients atteints d'un cancer de l'ovaire précoce. Les p-valeurs

indiquées sont asymptotiques, c'est-à-dire fondées sur l'hypothèse d'indépendance vis-à-vis des données. Afin de vérifier cette indépendance, des p-valeurs dites *empirical* sont générées correspondant à des p-valeurs virtuelles obtenues en permutant (n=10.000 fois) de manière aléatoire les cas patients et contrôles (*bootstrap*).

	1000G controls (n=154)		EVS controls (n=4300)	
	p-value	Empirical p-value	p-value	Empirical p-value
CAST	0.0025	-	0.0065	-
CMC	0.0043	0.0031	0.0050	0.0100
WSS	0.0012	0.0029	0.0011	0.0095
C-ALPHA	<0.0001	0.0017	0.0129	0.0188
SKAT	0.0016	0.0030	0.0204	0.0167

Tableau 8 : Tests multiples évaluant l'hypothèse d'un enrichissement en variants rares chez les patients

La seconde approche permettant d'appuyer la pathogénicité de cette variation *de novo* dans la survenue d'un cancer de l'ovaire précoce repose sur des données fonctionnelles. Nous avons évalué *in vitro* la fonctionnalité des protéines mutantes, en mesurant l'activité des protéines Activine A et Inhibines A et B dans des lignées cellulaires HEK-293F. Le vecteur plasmidique pcDNA3.1 (Invitrogen) comportant les ADN complémentaires humains *full lenght INHBA* ou *INHA* a été soumis à une mutagenèse dirigée (*Quick Change Lightning Mutagenesis Kit de Stratagene*) et transfecté par Lipofectamine (Lipofectamine 2000, *Life Technologies*). L'activité protéique a été mesurée par test immunologique, respectivement par E4/R1 et C5/R1 pour les Inhibines A et B, et E4/E4 pour l'Activine A [262]. La variation p.Asn386Ser d'*INHBA* induit dans ce système une réduction significative de 28% de la production d'Activine A ($p=0,0007$) et une augmentation de 44% également significative ($p=0,048$) de la production d'Inhibine A (Figure 6, *two-tailed t-test*). De manière encore plus drastique, le mutant *INHA* p.Arg60Leu effondre la production d'Inhibine A (-75%, $p=0,00006$) et d'Inhibine B (-55%, $p=0,002$). Cet effet majeur peut s'interpréter comme consécutif à la perte du domaine de clivage furine-dépendant du prodomaine, induit par cette variation. Comme indiqué précédemment, le ratio Inhibine/Activine apparaît comme impliqué dans la régulation endocrine et surtout paracrine de l'épithélium ovarien, l'Activine agissant comme inhibiteur de la prolifération épithéliale. De même les souris *INHA*^{-/-} tendent à développer des tumeurs gonadiques. Bien que ces observations ne constituent pas de démonstration formelle de l'induction d'une carcinogenèse ovarienne par ces mutants, l'effet protéique observé apparaît cohérent avec le développement de tumeurs épithéliales ovariennes.

Au terme de ce premier trio d'exomes soustractifs avec identification d'une variation délétère *de novo* du gène *INHBA*, impliquant une voie biologique identifiée comme participant à la carcinogenèse ovarienne, avec données fonctionnelles supportant un effet protéique, avec

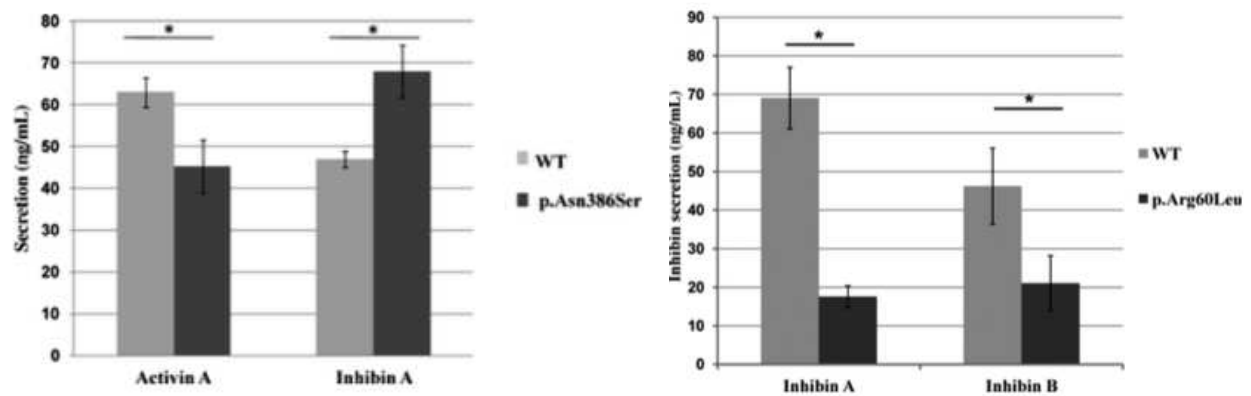


Figure 12 : Données d'activité par méthode immunologique des protéines Activine et Inhibine en situation wild-type et mutée

données de récurrence dans des cohortes de phénotype similaire, nous avons publié ces résultats dans la revue *Human Mutation* en 2014. Ce premier trio étant déjà initié lors de mon arrivée au Laboratoire, ma contribution à ce projet a concerné la recherche de récurrence par le séquençage Sanger des gènes *INHBA* et *INHA* des cohortes de cancers de l'ovaire, ainsi que la synthèse des données fonctionnelles, des données statistiques de recherche d'enrichissement en variants rares des gènes *INHBA* et *INHA*, et la participation à la rédaction de l'article.

Article : **Germline mutations of inhibins in early-onset ovarian epithelial tumors.**

Tournier I, Marlin R, Walton K, Charbonnier F, Coutant S, **Théry JC**, Charbonnier C, Spurrell C, Vezain M, Ippolito L, Bougeard G, Roman H, Tinat J, Sabourin JC, Stoppa- Lyonnet D, Caron O, Bressac-dePaillerets B, Vaur D, King MC, Harrison C, Frebourg T.

Hum Mutat. 2014 Mar;35(3):294-7.

Exomes soustractifs dans un trio parent / enfant affecté par un cancer du sein précoce.

Contexte Clinique

Encouragés par l'identification d'un variant *de novo* pathogène dans une voie biologique impliquée dans le phénotype tumoral du cas index, nous avons tenté de reproduire cette méthodologie des exomes soustractifs avec un second trio. Le cas index est ici une jeune femme de 21 ans affectée par un cancer du sein localement avancé, récepteurs hormonaux négatifs HER2 positif, classé cT4N1, traitée par chimiothérapie néo-adjuvante anthracyclines/taxanes + trastuzumab, mastectomie curage axillaire et radiothérapie. Les deux parents ainsi que les autres apparentés ne présentaient pas d'historique néoplasique particulier. Les analyses *BRCA1*, *BRCA2* et *TP53* se sont avérées sans variation identifiée. Ma contribution à cet exome est ici complète, à savoir sélection du cas, réalisation des librairies d'ADN et séquençage, exploitation des résultats.

Matériel et Méthode

Nous disposions pour cette étude des ADN parentaux extraits d'ADN leucocytaire, de l'ADN leucocytaire du cas index, ainsi que de l'ADN tumoral et tissulaire cutané normal (derme mammaire) pour le cas index. Les exomes ont été réalisés à partir de librairies d'ADN préparées selon la méthode QXT (*SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing*). Cette méthode diffère de la technologie XT par la technique de fragmentation de l'ADN, non plus mécanique par sonication mais enzymatique par transposase. La réparation des extrémités et la ligation des adaptateurs est également effectuée dans une seule réaction lors de la fragmentation par la transposase. La quantité d'ADN initiale (*input*) est également réduite dans cette technologie, correspondant à 100 ng au lieu des 3 µg de la technologie XT. Les étapes suivantes sont en revanche similaires, soient les étapes de purification sur billes AMPures XP, l'amplification pré-capture par PCR 8 cycles, le contrôle de la qualité des fragments d'ADN sur TapeStation. La capture des exomes a été de manière identique effectuée par le kit *SureSelect Human All Exons 50 Mb kit* (24000 genes, Agilent Technologies). Le séquençage a été réalisé sur la plateforme GAIIx. Le pipeline bioinformatique est également similaire avec une analyse primaire effectuée par le logiciel RTA 1.10 d'Illumina, le démultiplexage, l'alignement et le *variant calling* par CASAVA (version 1.8.2 et génome de référence Human hg19/GRCh37), l'annotation des variants par le double pipeline *ANNOVAR* et *Variant Effect Predictor*. La filtration des variants obtenus et la soustraction des variants d'origine parentale à l'exome du cas index ont également été réalisées à l'aide du logiciel EVA.

Résultats

La filiation du cas index a été vérifiée avec succès par l'étude microsatellitaire.

Nous avons obtenu en moyenne 6,6 Mb par exomes avec une profondeur moyenne de 76x, et 96% des bases lues à une profondeur >10x. Les exomes sont de haute qualité, avec un alignement correct pour 97% des *reads* et en moyenne 92% des bases présentant un Q-score supérieur à 30. Nous avons appliqué les mêmes règles de filtration assistée du logiciel EVA et identifié deux variations *de novo* confirmées sur les BAM et en Sanger, présentes chez le cas index et absentes chez les parents.

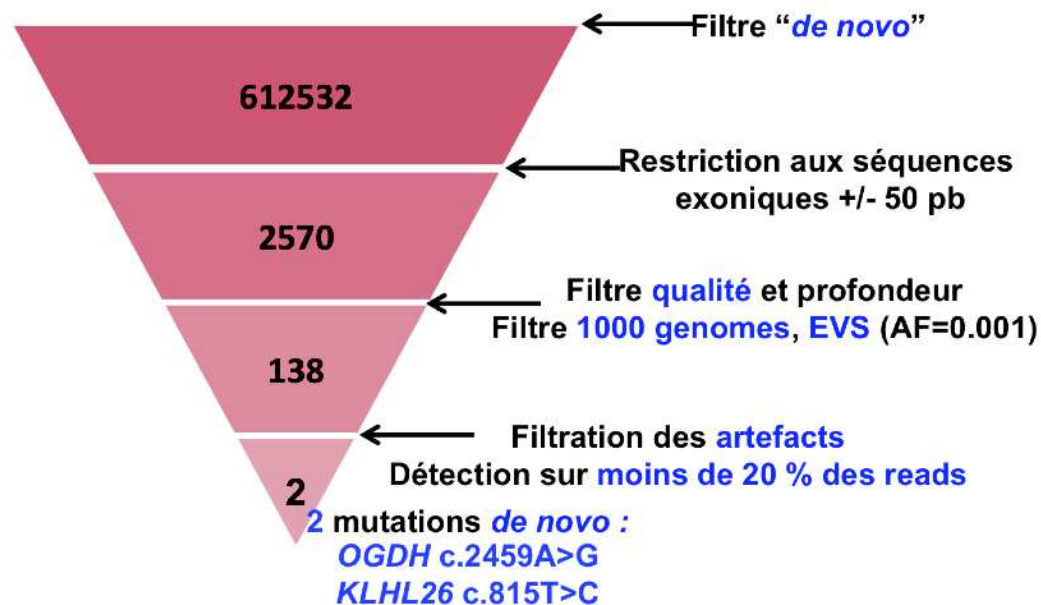


Figure 13 : Filtration assistée du logiciel EVA

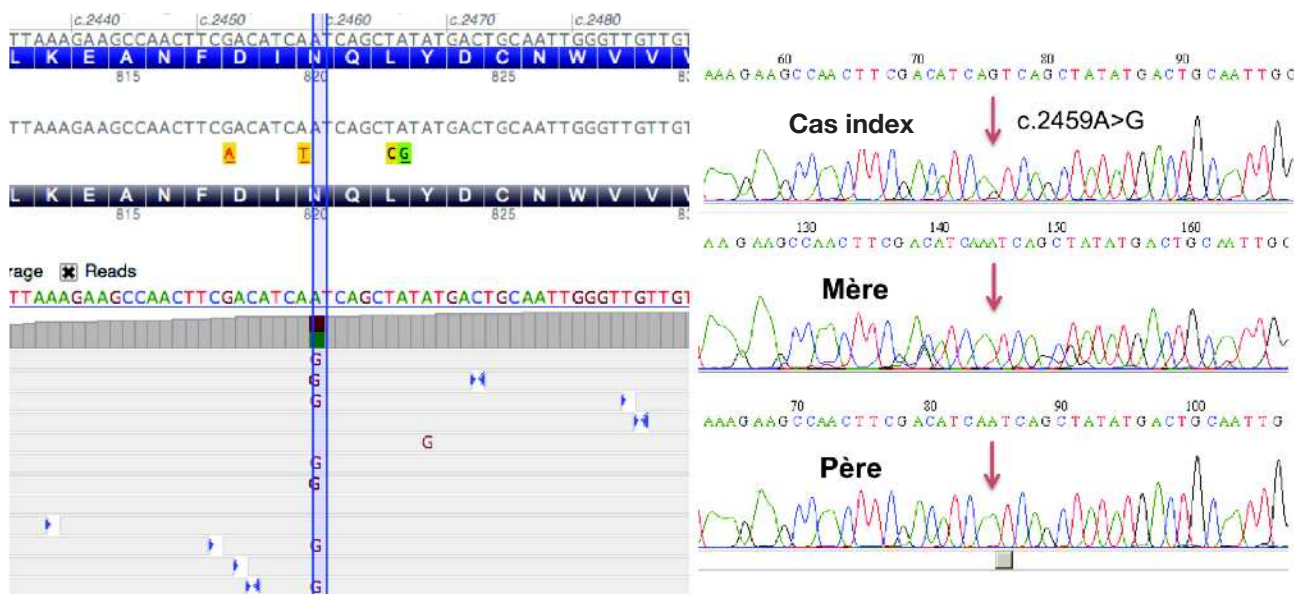


Figure 14 : BAM et séquences Sanger correspondant à la variation c.2459A>G d'OGDH

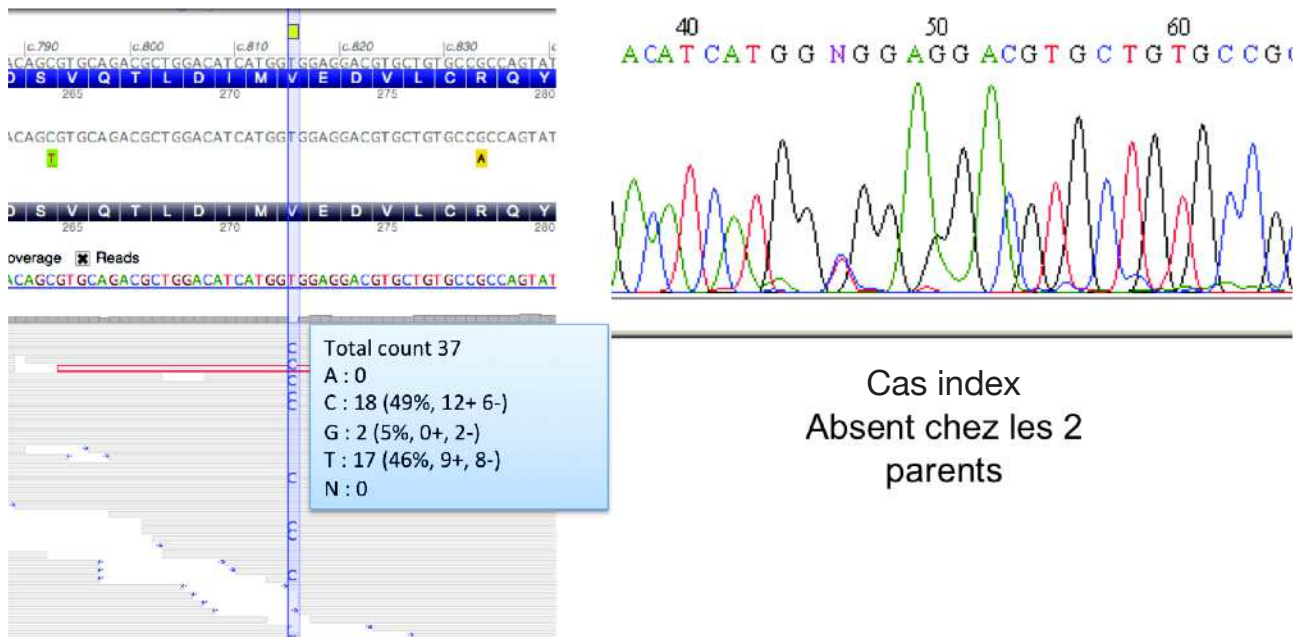


Figure 15 : BAM et séquence Sanger correspondant à la variation c.815T>C de *KLHL26*

La variation observée sur le gène *OGDH* (MIM*613022) est une variation faux-sens intéressant l'exon 20, parfaitement hétérozygote avec une balance allélique de 50% sur les BAM et une profondeur de lecture importante (309x), transformant le résidu Asparagine 820 en Sérine. La distance de Grantham pour cette substitution est de 46. Cette variation est répertoriée dans les bases de données, avec le référencement rs375313423 dans dbSNP, et des fréquences rapportées dans la base ExaC de 9.061e-05 en population générale, principalement portée par les populations asiatiques. Le gène *OGDH* semble peu tolérant aux variations faux-sens, avec un z-score de 5,19, et également très intolérant aux variations pertes de fonction avec un *pLI*-score de 0,9. Les prédictions de pathogénicité ne sont pas en faveur d'un effet délétère, que ce soit pour AlignGV/GD (classe C0), PolyPhen-2, MutationTaster et SIFT. Il n'est pas prédit non plus d'effet sur l'épissage. Bien que la position 820 soit impliquée dans les domaines protéiques *transketotase-like pyrimidine-binding* et *2-oxoglutarate dehydrogenase - E1 component*, la conservation protéique à cette position est faible jusqu'au rat. L'étude fonctionnelle du pouvoir catalytique de *OGDH* chez le rat (avec le résidu Serine en position 820) et chez le porc (avec le résidu Asparagine en position 820) ne montre pas de différence significative (*données personnelles de Richard Denton*) [263]. Ainsi, bien que le gène *OGDH* apparaisse comme un candidat pertinent par son implication forte dans le cycle de Krebs et les aspects connus de la carcinogénèse induite par des onco-métabolites (variants d'*IDH* dans les leucémies aigues myéloïdes et les glioblastomes secondaires, variants de *FH* dans la léiomyomatose familiale, variants de *SDH* dans les paragangliomes), cette **variation apparaît neutre et sans relation avec le phénotype tumoral extrême observé.**

La variation observée c.815T>C sur le gène *KLHL26* (Entrez Gene 55295) est également une variation faux-sens p.(Val272Ala), hétérozygote sur les BAM (49/46%) mais avec une faible profondeur de lecture (37x) néanmoins confirmée en séquençage Sanger. La distance de Grantham de cette substitution apparaît moyenne (64), et référencée dans *ExaC* avec une MAF de 8.606e-06 portée exclusivement par la population africaine, et un référencement dans dbSNP (rs375188353). Le gène *KHLH26* apparaît peu contraint, avec un z-score de 3,19 et un *pLI*-score de 0. Les prédictions de pathogénicité pour la variation c.815C>T apparaissent néanmoins fortes, avec une classe C25 pour AlignGVGD, le statut *deleterious* pour SIFT et *disease-causing* pour MutationTaster, en revanche le statut *Benign* pour PolyPhen-2. Les prédictions d'épissage ne sont pas en faveur d'un effet délétère, en dehors du renforcement observé de l'*ESE SF2/ASF* prédit par *ESE Finder*. La variation affecte le domaine protéique *gigaxonin*, avec une faible conservation protéique à cette position (Leucine chez le rat et la souris). *KLHL26* appartient à une super-famille composée de 42 membres ayant en commun les domaines *BTB/POZ*, *BACK* et plusieurs motifs *KELCH*. Les fonctions protéiques semblent diverses et mal connues, impliquant les échanges ioniques, le cytosquelette, la dégradation protéique via l'ubiquitination, la régulation transcriptionnelle [264]. Des variations impliquant des membres de cette super famille apparaissent impliquées dans des pathologies diverses, allant de l'hypertension artérielle familiale (pseudohypoaldostérisme type IID (OMIM: 614495)) pour des variants de *KLHL3* à la rétinite pigmentaire autosomique dominante pour *KLHL7* (MIM: 612943), en passant par la neuropathie géante axonale (MIM: 256850) pour les variants de *KLHL16*. Des altérations somatiques de *KLHL6* ont été identifiées récurrentes dans la leucémie lymphocytaire chronique, des variants pertes de fonction de *KEAP1* (*KLHL19*) sont associés à la progression tumorale dans de nombreux types tumoraux et des variants somatiques de *ENC1* (*KLHL37*) sont associés à des tumeurs cérébrales. En dépit de ces observations, la fonction peu connue de *KLHL26* et de son rôle potentiel dans la cancérogénèse, les prédictions discordantes, la faible conservation protéique nous ont conduit à ne pas retenir davantage cette variation *de novo* comme causale dans le phénotype observée chez le cas index.

Nous avons également considéré ce trio d'exomes sous l'hypothèse d'un modèle récessif. Dans une première approche, nous avons recherché des altérations homozygotes chez le cas index et hétérozygotes chez les deux parents. Afin de réduire le nombre de variants, une filtration a été appliquée afin de retenir les variants concernant les régions codantes exoniques et les séquences introniques bordantes (+/- 10 paires de bases), non synonymes, présentant des données de qualité (Q-score>30) et de couverture (>20x) et enfin rares (MAF ESP et 1000 *Genome* <0,01). Les statuts homozygotes et hétérozygotes ont été vérifiés sur les BAM des exomes du cas index et des parents. Aucune variation répondant à ces restrictions et présentant des prédictions bio informatiques en faveur d'un caractère délétère n'a été détectée. Dans un deuxième temps, nous avons considéré un modèle hétérozygote composite qui suppose que le cas index ait hérité de deux allèles mutés différents de chacun de ces parents, au sein du même gène. Nous avons de ce fait appliqué une filtration afin de sélectionner les variants faux-sens ou

Exomes comparatifs dans une famille lourdement affectée par des cancers du sein ou du pancréas précoces

Le Service de Génétique Clinique du CHU et le Centre Henri Becquerel de Rouen ont pris en charge à des époques différentes une famille fortement évocatrice d'un risque génétique particulier associant cancers du sein et cancers du pancréas précoces. Plusieurs cas index avaient été analysés (III.3, III.16, III.18 et IV.1) sans altération identifiée sur les gènes canoniques de la prédisposition héréditaire au cancer du sein et de l'ovaire. Cette famille est en effet remarquable par les âges d'apparition du premier cancer du sein (33 ans, 23 ans, 37 ans et 27 ans respectivement), la sur-expression de la protéine HER2 chez 3 de ces jeunes femmes, le caractère bilatéral du cas III.18, l'association sur 4 générations de cancers du sein et de cancers du pancréas. L'arbre de cette famille remarquable est schématisé dans la figure 10.



Nous avons appliqué à cette famille fortement évocatrice une stratégie d'exomes comparatifs, disposant des ADN leucocytaires des cas III.3, III.16, III.18 et IV.1. L'intérêt de cette approche ici présente est la distance méiotique séparant chacun de ces individus, s'agissant de cousines au 1er degré ou d'une nièce (distance méiotique entre 3 et 4). Ainsi, la présence de variations communes aux quatre exomes des personnes atteintes constituerait un argument de pathogénicité, compte tenu du nombre d'individus comparés et des distances génétiques les séparant. L'ensemble de cette partie a de nouveau été entièrement réalisée par mes soins.

Matériel et Méthode

La préparation des librairies d'ADN a été effectuée pour cette étude selon la méthodologie QXT (*SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing*). Les exomes ont été capturés selon le kit *SureSelect Human All Exons 50 Mb kit* (24000 genes, Agilent Technologies). Le séquençage a été réalisé sur la plateforme Illumina NextSeq™ 500. Le pipeline bioinformatique est similaire aux exomes précédemment réalisés. Pour cette stratégie de recherche de récurrence, un fichier Excel regroupant l'ensemble des variants détectés chez les quatre individus a été généré et travaillé, afin d'effectuer les filtrations sur les critères de qualité et de pathogénicité. Les variants ont été contrôlés à partir des fichiers BAM sur l'interface Alamut (Interactive Biosoftware).

Résultats

Les exomes séquencés ont présenté de très bons paramètres de qualité, avec en moyenne 10Mb par exomes, plus de 95% des bases présentant un Q-score supérieur à 30 et un Q-score moyen de 34,5. L'alignement est également correct avec en moyenne 97,5% des *reads* correctement alignés. La profondeur moyenne est autour de 100x, avec plus de 95% des *reads* séquencés à une profondeur supérieure à 30x.

L'analyse combinée des 4 exomes identifie 10088 variants exoniques +/- 10 pb avant toute filtration. L'objectif étant l'identification de variations exoniques ou d'épissage prédites délétères et communes à au moins 3 des 4 individus apparentés, l'algorithme de filtration suivant a été appliqué :

- Filtration sur la qualité : Q-score >30 et profondeur minimale supérieure à 30x : 8196 variants résiduels
- Filtration sur la conséquence de la variation au niveau génique avec exclusion des variations synonymes ou des *indels* en phase : 7577 variants résiduels
- Filtration sur la récurrence avec présence de la variation chez au moins 3 des apparentés : 625 variants résiduels

- Filtration sur la balance allélique recherchant des variations hétérozygotes : 103 variants résiduels en retenant une balance allélique supérieure incluse dans le ratio 30/70
- Filtration sur la fréquence ExaC en retenant les variations rares inférieures à 0,001 : 50 variants résiduels.
- Filtration sur les algorithmes de prédiction en ne retenant que les variants prédits délétères par au moins un algorithme parmi SIFT, Polyphen-2 et MutationTaster : 41 variants résiduels.

Les variations retenues ont été visualisées sur les BAM des 4 individus séquencés et s'avèrent artéfactuelles ou correspondre à des erreurs d'annotations. **Il n'est donc pas retenu au terme de cette analyse de variations ponctuelles exoniques prédites délétères partagées par au moins 3 individus de cette famille remarquable.**

La recherche de réarrangements de grande taille (CNV) éventuellement communs à ces individus n'a pu être effectuée, l'outil CANOES nécessitant un nombre supérieur d'échantillon afin d'être en capacité de détection.

Discussion concernant les stratégies d'exomes soustractifs / comparatifs

Au terme de cette première approche, nous observons plusieurs niveaux de complexité dans la recherche de variations délétères par séquençage d'exomes. L'hypothèse de variations *de novo* délétères apparues chez des individus atteints de phénotype extrême sous la forme de cancer précoces et agressifs justifiait l'approche par exome soustractif. Sur les deux trios présentés ici, auquel s'associe un troisième trio effectué au sein de l'unité portant sur un second cas de cancer du sein précoce avec identification de deux variations *de novo* affectant *VAR2* et *FAT1* (Thèse de Mme MARLIN Régine), nous n'avons pu faire la démonstration de l'implication causale de la variation *de novo* uniquement dans le contexte du cancer de l'ovaire précoce, avec variant *de novo* de *INHBA*. Dans ce cas, la voie biologique impliquée, la récurrence observée dans une cohorte de cancers de l'ovaire précoces, les données fonctionnelles en faveur d'une altération de la balance Activine/Inhibine nous ont permis d'imputer cette variation comme causale. Le gène *INHBA* a depuis été intégré dans le panel de gène analysé à titre exploratoire par la plate-forme diagnostic du Centre F. Baclesse (D. Vaur) sans élément de récurrence cependant observé à ce jour. A l'opposé, dans les deux trios correspondant à des jeunes femmes affectées par des cancers du sein précoce, les variations *de novo* mises en évidence (*VAR2* et *FAT1*, *OGDH* et *KLHL26*) ne présentaient pas l'ensemble des caractéristiques permettant de conclure à leur imputabilité : absence de récurrence et données fonctionnelles peu robustes pour *FAT1*, substitution a priori tolérée pour *OGDH*, prédictions discordantes pour *KLHL26* et un gène dont la fonction est peu connue et sans relation évidente avec le phénotype. L'une des différences fondamentales entre ces trios pouvant contribuer à ces résultats divergents est l'extrême rareté d'un cancer de l'ovaire aussi précoce et dans une forme de bas grade également plus rare. A l'inverse, la survenue d'un cancer du sein avant 31 ans constitue en proportion un

évènement plus fréquent, pour lequel un déterminisme génétique doit être évoqué mais sans exclure un mécanisme accidentel somatique.

Les exomes comparatifs dans cette famille remarquable de femmes atteintes précocement (avant 40 ans voire avant 30 ans) de cancers du sein agressifs, HER2 surexprimé, associé à un cancer du pancréas précoce, n'ont pas permis d'identifier une altération commune. Notre approche s'apparentait aux stratégies antérieures des études de liaison, avec cependant une puissance ici limitée par le nombre d'individus. Si un déterminisme génétique est ici fortement probable aux vues de cet arbre, une altération cryptique doit être évoquée, notamment au niveau de *BRCA1* ou *BRCA2* devant les phénotypes observés.

Ainsi, cette première approche a permis d'illustrer la survenue de variations *de novo* dans la proportion attendue (0-3 variations par exome) mais l'interprétation de l'effet de ces variants et de leur implication dans les phénotypes observés s'est heurtée à la difficulté d'accessibilité à des tests fonctionnels validés, s'agissant de gènes parfois peu étudiés, ainsi qu'à la difficulté d'effectuer des études de récurrence au sein de grandes cohortes, pour des situations par définition rares compte tenu du caractère phénotype extrême. Cet argument de causalité par démonstration de la récurrence est également en contradiction avec le concept de variations *de novo*, corollaire du concept de variation privée. D'autre part, l'accessibilité permise par le NGS à la complexité du génome humain a souligné la très grande fréquence des SNV (autour de 20.000 par exome) [265] et dont la très grande majorité est neutre. Ainsi, dans une étude par NGS de 114 gènes chez 1000 personnes issues du programme ESP, 239 variants ont été identifiés comme délétères dans HGMD. L'étude approfondie des éléments supportant ce caractère délétère a permis de réduire ce nombre de variants possiblement pathogènes au nombre de 16, soit 6% des variants rapportés [266]. Ces limites expliquent probablement l'asymétrie du nombre de publications entre les situations cliniques correspondant aux anomalies du développement (1648 articles utilisant le WES dans ces conditions) et les situations correspondant aux cancers héréditaires (160 articles uniquement).

Notre conclusion à ce stade, devant les difficultés pour apporter les éléments d'imputabilité pour ces variants *de novo*, a été de considérer une liste restreinte de gènes, sélectionnés pour leur implication évidente dans la carcinogenèse. De ce fait, les altérations identifiées pourront être présumées causales. Nous avons appliqué cette stratégie dans différentes cohortes de patients répondant à notre définition de phénotype extrême, soit cancer de précocité et/ou d'agressivité inhabituelle.

Partie II : séquençage par panel de gènes impliqués dans la cancérogenèse de cohortes d'individus affectés par des tumeurs de phénotype extrême

Rationnel et Contexte

Cette seconde approche s'est nourrie de la publication fondamentale du *Broad Institute* [267], à l'origine de la mise en évidence d'un nombre restreint de gènes impliqués dans la cancérogenèse. En effet, si la description des événements génétiques dans les tumeurs concerne un grand nombre de gènes, tel que rapporté dans *The Cancer Genome Atlas* (TCGA) ou le *Catalogue Of Somatic Mutations In Cancer* (COSMIC), la plupart de ces événements concernent des variations passagères. Si la notion de gènes suppresseurs de tumeurs et d'oncogènes est connue depuis le début du XX^{ème} siècle, il apparaît au travers de cette publication qu'une liste limitée à quelques centaines de gènes est retrouvée de manière récurrente dans la quasi intégralité des types tumoraux, correspondant à des variations dites *drivers*, c'est-à-dire conférant un avantage sélectif à la cellule tumorale à l'origine de sa prolifération et de ses capacités invasives. Ces notions de variations *drivers* et *passengers* sont déjà définies dans la publication de Vogelstein [268] où un nombre d'environ 140 gènes sont identifiés comme récurrents et à l'origine de variations initiatrices. Les dissections moléculaires des types tumoraux largement publiées ces dernières années sous les termes de « *genomic landscape of* » ou « *molecular portrait of* » rapportent ainsi pour les différents types tumoraux diverses variations somatiques, affectant en moyenne autour d'une cinquantaine de gènes. Une diversité de cette charge mutationnelle est néanmoins rapportée en relation avec l'exposition aux carcinogènes (UV et mélanome) et/ou des altérations de la voie de la réparation de l'ADN (*POLE*) [269]. Ainsi, un classement des tumeurs en fonction de la charge mutationnelle peut être proposé, à l'origine notamment d'une compréhension de l'efficacité de l'immunothérapie pour les tumeurs de type mélanome, cancer du poumon ou tumeurs avec profil MSI en raison de leurs charges mutationnelles élevées produisant des néo-antigènes [270]. Au-delà de ces descriptions globales, l'identification des altérations *drivers* et de la séquence de ces altérations à l'origine de la transformation maligne sont critiques, dans une perspective théranostique. Ainsi, le modèle très connu de la transformation des adénomes en adénocarcinomes lieberkühniens illustre parfaitement ce concept : altération d'*APC* constituant la variation fondatrice ou *gatekeeping*, suivie d'une altération de *KRAS* ouvrant la voie à la prolifération de ces cellules et enfin altérations de *TP53*, *SMAD4* et *PI3KCA* conférant aux cellules tumorales les avantages sélectifs permettant le développement d'un cancer infiltrant. Une méthode d'identification de ces altérations *drivers* est proposée par Vogelstein *et al* [268], selon le type de variations et leur caractère récurrent (*hotspot*). Les variations *drivers* des oncogènes sont des variations faux-sens, et les altérations *drivers* des gènes suppresseurs de tumeurs sont de type tronquantes. Un ratio arbitraire de 20/20 est proposé pour définir une altération *driver*, dès lors que plus de 20% des variations rapportées

pour ce gène sont à position constante et de type faux-sens pour les oncogènes, ou que plus de 20% des variations de ce gènes sont de type tronquantes pour les gènes suppresseurs de tumeurs. Sur cette base de définition, Vogelstein *et al* propose ainsi une liste de 125 gènes concernés par des altérations *drivers*, auxquels s'ajoutent 13 autres gènes affectés par des événements de type translocations ou réarrangements de grande taille récurrents et *drivers*. Lawrence *et al* [260] ont recherché spécifiquement ces altérations *drivers* (appelées ici *candidate cancer genes*) par un séquençage massif de 4742 paires tissu tumoral (TT) / tissu normal (TN) correspondant à plus de 21 type tumoraux différents. Ce séquençage identifie en moyenne 672 altérations par paire TT/TN, correspondant à des SNV ou *small indels*. Les *candidate cancer genes* sont définis comme : (i) enrichis en variations comparativement à la mutabilité attendue (z-score), (ii) affectés à des positions récurrentes type *hotspots*, (iii) enrichis en variations affectant des positions conservées. Différentes listes sont ainsi proposées pour chaque type tumoral, ou pour la cohorte globale. Une liste de 254 gènes appelée *Cancer5000* est ainsi proposée et accessible sur le site <http://www.tumorportal.org/>. Par une analyse de saturation, les auteurs montrent en outre que pour les gènes affectés de manière récurrente (>20% des tumeurs), la liste de gènes n'est que marginalement affectée par le nombre d'échantillons analysés. Cette observation pose de ce fait le constat d'une liste restreinte d'environ 250 gènes affectés de manière récurrente et *drivers* dans la plupart des types tumoraux.

La seconde observation est relative à la superposition entre les gènes affectés par des variations constitutionnelles à l'origine des syndromes de prédisposition héréditaire, et les gènes appartenant à cette liste de *candidate cancer genes*. Il en est ainsi pour *TP53* altéré dans plus de 80% des types tumoraux et à l'origine du syndrome de Li-Fraumeni, *BRCA1* et *BRCA2* à l'origine de la prédisposition héréditaire au cancer du sein et de l'ovaire mais aussi altérés uniquement en somatique dans des cancers de l'ovaire ou du sein sporadiques, *PTEN* à l'origine du syndrome de Cowden mais altéré également dans le cancer du sein, *RET* à l'origine de la néoplasie endocrinienne multiple de type II mais aussi partenaire de translocations causales dans les cancers folliculaires de la thyroïde [271], les gènes de la voie RAS/MAPK (*KRAS*, *HRAS*, *NRAS*, *NF1*...) altérés dans de nombreux cancers (digestif, pulmonaire, tête et cou...) et également à l'origine des syndromes poly-malformatifs et prédisposant aux cancers que représentent les *rasopathies*. Nous pouvons dès lors faire l'hypothèse que des altérations constitutionnelles sur des gènes appartenant à la liste de *candidate cancer genes* peuvent être à l'origine d'une prédisposition héréditaire aux cancers, et dès lors que cette liste de 254 gènes représente également une liste de *candidate cancer predisposition genes*.

Un autre élément remarquable de l'étude de Lawrence *et al* est l'observation d'un nombre, certes restreints, de *candidate cancer genes* fréquemment altérés **quel que soit le type tumoral**. Ainsi, *TP53*, *PIK3CA*, *PTEN*, *RB1*, *KRAS*, *NRAS*, *BRAF*, *CDKN2A*, *FBXW7*, *ARID1A* et *MLL2*, *STAG2* se retrouvent altérés dans 4 ou plus types tumoraux différents [267]. Additionnellement, *ATM*, *CASP8*, *CTCF*, *ERBB3*, *HLA-A*, *HRAS*, *IDH1*, *NF1*, *NFE2L2* et *PIK3R1* sont altérés dans trois types tumoraux différents. Ainsi, si un certain nombre d'altérations somatiques peuvent constituer

une origine commune à différentes cancérogenèses, et si ces altérations, lorsque présentes à l'état constitutionnel, constituent en outre un risque héréditaire de cancers, nous pouvons supposer qu'une altération constitutionnelle d'un des ces gènes peut être à l'origine d'un spectre de prédisposition plus large que ne prévoyait les descriptions originelles. Ce **concept d'extension phénotypique** est vérifié dans la littérature, avec pour preuves les définitions élargies du syndrome de Li-Fraumeni [98, 99], la survenue significativement plus élevée qu'en population générale de cancers hors spectre chez les porteurs de variants *BRCA1* et *BRCA2* (colon, mélanome, voies biliaires, estomac) [35, 272], ou respectivement un sur-risque de cancer du sein chez des porteurs d'altération des gènes *MMR* [198]. Cette hypothèse d'extension phénotypique justifie également, certes de manière discutée, la réalisation d'analyses en panel de gènes élargis pour le diagnostic de formes héréditaires de cancer.

Enfin, les événements affectant ces gènes peuvent survenir en **mosaïque**, c'est-à-dire n'affectant qu'une partie des tissus de l'organisme dès lors que l'altération est survenue *de novo* en post-zygotique. Dans ces situations par définition sporadiques (parents non porteurs), l'altération peut ne pas être détectable par les techniques de séquençage standard si elle ne concerne pas le tissu hématopoïétique, sauf à utiliser une technique à forte profondeur. Peu de données sont disponibles dans la littérature concernant ces événements en mosaïque, principalement par défaut de profondeur de séquençage. Leur réalité est cependant attestée lors de la conception de cette sous-étude par deux publications d'une altération de *TP53* en mosaïque [273, 274]. De même, une altération constitutionnelle de *BRCA1* présente sous forme de mosaïque chez une jeune femme affectée de cancer du sein triple négatif est rapporté en 2015 [275]. Nous avons donc considéré cette hypothèse pour calibrer le nombre d'échantillons testés par *run* de séquençage afin de disposer d'une profondeur suffisante permettant la détection de ces événements.

Ainsi, nous avons réalisé le **séquençage à grande profondeur de cohortes d'individus affectés par des cancers de phénotype extrême, négatifs pour les analyses constitutionnelles canoniques, en définissant une capture restreinte à un ensemble de gènes impliqués dans la cancérogenèse et ainsi candidats à un déterminisme héréditaire.**

Patients et Méthodes

Patients

Nous nous sommes appuyés sur plusieurs collections biologiques annotées sur le plan clinique et sur le recrutement prospectif des unités de génétique clinique affiliées à l'unité Inserm 1245 ou sollicitant le Laboratoire de Génétique de Rouen pour son expertise dans les domaines des prédispositions héréditaires au cancer colorectal et du syndrome de Li-Fraumeni. Nous avons ainsi différencié 4 cohortes de patients selon la pathologie présentée :

- Cancer du sein précoce défini par la survenue d'un cancer infiltrant avant 31 ans (THERY JC - Rouen, MULLER Etienne - Centre François Baclesse, Caen)

- Cancer de l'ovaire précoce défini par la survenue d'un cancer de l'ovaire avant 41 ans (MULLER Etienne - Centre François Baclesse, Caen)
- Cancer colorectal précoce défini par la survenue d'un cancer colorectal avant 31 ans (CHARBONNIER Françoise - Rouen)
- Cancers pédiatriques de présentation extrême : âge précoce inhabituel, tumeur rare ou tumeurs multiples (FERMEY Pierre, RENAUX-PETEL Mariette - Rouen).

Les différentes cohortes ont été recrutées, séquencées et analysées par ces différentes parties prenantes de cette large étude, toutes membres de l'unité Inserm 1245 (site de Rouen, Faculté de Médecine Pharmacie - site de Caen, Centre François Baclesse). Etant personnellement concerné uniquement par la cohorte des cancers du sein précoce, l'exposé des résultats se limitera à cette cohorte. Néanmoins, l'interprétation finale des résultats sera pour partie globalisée afin de rechercher des enrichissements en variants rares communs aux différentes cohortes.

Le recrutement des cas étudiés affectés d'un cancer du sein précoce s'est effectué à partir des patients vus en Unité de Génétique Clinique à Rouen, de manière prospective mais aussi rétrospective dès lors qu'un échantillon d'ADN de qualité était disponible au Laboratoire. La collection biologique des ADN des patients analysés au sein du Laboratoire de Génétique de Rouen pour la recherche d'une altération *TP53* dans le contexte d'un cancer du sein avant 31 ans a également été filtrée, sur la négativité des analyses *TP53* et *BRCA1/2*, et de *PALB2* lorsqu'effectuée (à partir de 2015). La recherche d'un consentement du patient autorisant la poursuite des analyses génétiques à titre exploratoire dans le contexte de sa pathologie conditionnait également l'utilisation de son ADN. Les prescripteurs initiaux de l'analyse *TP53* justifiant de la présence de l'ADN au sein du Laboratoire ont été avertis de cette étude. Lorsque les informations étaient disponibles, la structure familiale et les antécédents familiaux de cancer ont été annotés. De même, le phénotype tumoral a été collecté dès lors que disponible, pour d'éventuelles corrélations génotype/ phénotype.

Nous avons par ailleurs constitué une cohorte de témoins afin de rechercher un éventuel enrichissement en variants rares parmi notre capture de gènes chez les patients. Nos cohortes étant composées de patients reflétant la diversité ethnique française, nous avons souhaité extrapoler cette cohorte de témoins à partir de données de séquençage d'individus français. Nous avons pu disposer à cette fin de la cohorte de témoins français du projet France Exome (FREX).

Définition de la capture de gènes

Comme indiqué précédemment, nous avons fait l'hypothèse que les gènes impliqués de manière récurrente dans la cancérogenèse pouvaient être également altérés de manière constitutionnelle dans des présentations cliniques exceptionnelles par leur sévérité ou précocité, et ainsi constituer les déterminants de formes héréditaires de cancers. La liste de gènes a été

constituée de manière collaborative avec deux autres étudiants en Thèse impliqués dans cette étude au sein de l'unité Inserm 1245 : Pierre Fermey et Etienne Muller. La méthodologie retenue repose sur une première étape réunissant l'ensemble des gènes impliqués dans les cancers par extraction des bases de données disponibles :

- Base COSMIC (<http://cancer.sanger.ac.uk/cosmic>)
- Portail du *Broad Institute* en relation avec la publication de Lawrence *et al* [267] (<http://www.tumorportal.org>)
- Portraits moléculaires des tumeurs étudiées, en particulier les données du TCGA (<http://www.cbioportal.org>) et autres portraits publiés [276-278]
- Panels académiques ou commerciaux : panel diagnostique de l'Inca relatif aux prédispositions héréditaires de cancers (<https://www.e-cancer.fr/Professionnels-de-sante/L-organisation-de-l-offre-de-soins/Oncogenetique-et-plateformes-de-genetique-moleculaire/Les-predispositions-genetiques>), panels Life, TruSeq, Quiagen, MSK-Impact, etc...
- Base OMIM filtrée sur les syndromes de prédisposition héréditaire aux cancers (<https://www.omim.org>)

L'ensemble de ces interrogations nous ont permis de retenir, après élimination des doublons, plus de 800 gènes impliqués dans la cancérogenèse. Les niveaux de preuve de ces implications étant hétérogènes, et notre volonté étant de restreindre cette liste de gènes à environ 200 gènes pour garantir une profondeur de séquençage importante, nous avons ensuite réalisé une cure manuelle par analyse des données publiées relatives à l'implication de ces gènes. Afin d'aider à la hiérarchisation de cette liste, nous avons élaboré ce score de pondération :

- Présence du gène dans les panels de gènes académiques ou commerciaux parmi Cancer-5000S, Cancer-5000 Ov, Cancer-5000 CCR, Panel HPV (défini par le Laboratoire de Caen pour la cancérogenèse du cancer du col), Panel HBOC du Laboratoire de Caen, Panel Inca de la prédisposition héréditaire au cancer, Panel *Life Comprehensive Cancer*, Panel *TruSeq Amplicon*, Panel Qiagen, en affectant 1 point par présence dans un des panels. Les panels académiques ont été valorisés d'un facteur 10 compte tenu de la typologie des cancers investigués dans cette étude et de leur conception *a priori* davantage robuste.
- Présence du gène dans les panels spécifiques des tumeurs pédiatriques afin d'enrichir spécifiquement cette liste dans cette catégorie
- Evidences bibliographiques référencées dans Pubmed d'implications des altérations de ce gène à l'état **somatique** dans les cancers : 0 pour aucune, 1 pour discutable et 2 pour convaincante augmenté d'un facteur 3
- Evidences bibliographiques référencées dans Pubmed d'implication des altérations de ce gène à l'état **constitutionnel** dans les cancers : 0 pour aucune, 1 pour discutable et 2 pour convaincante, augmenté d'un facteur 10
- Le gène figure dans les 100 premiers gènes altérés dans la base COSMIC (*Top100 List*) : 0 ou 1, augmenté d'un facteur 3
- Le gène est impliqué dans les voies biologiques impliqués dans la cancérogenèse : 0 ou 1, augmenté d'un facteur 3

- Le gène est présent dans la base OMIM et associé à un risque de cancer : 0 ou 1, augmenté d'un facteur 10

Cette pondération nous a permis ainsi de définir une liste hiérarchisée de ces gènes. Nous avons procédé à différents arbitrages et compromis afin d'obtenir une liste représentative de l'ensemble des gènes impliqués ou candidats aux différentes cohortes de patients.

<i>AIP</i>	<i>EPHB2</i>	<i>KIT</i>	<i>RAD21</i>
<i>AKT1</i>	<i>ERBB2</i>	<i>KMT2B</i>	<i>RAD50</i>
<i>AKT2</i>	<i>ERCC1</i>	<i>KMT2C</i>	<i>RAD51</i>
<i>AKT3</i>	<i>ERCC2</i>	<i>KRAS</i>	<i>RAD51B</i>
<i>ALK</i>	<i>ERCC3</i>	<i>MALT1</i>	<i>RAD51C</i>
<i>AMER1</i>	<i>ERCC4</i>	<i>MAP3K1</i>	<i>RAD51D</i>
<i>APC</i>	<i>ERCC5</i>	<i>MAX</i>	<i>RAF1</i>
<i>ARAF</i>	<i>EXT1</i>	<i>MDM2</i>	<i>RB1</i>
<i>ARID1A</i>	<i>EXT2</i>	<i>MDM4</i>	<i>RECQL4</i>
<i>ARID5B</i>	<i>FAM157A</i>	<i>MEN1</i>	<i>RET</i>
<i>ATM</i>	<i>FANCA</i>	<i>MET</i>	<i>RINT1</i>
<i>ATR</i>	<i>FANCB</i>	<i>MGMT</i>	<i>RNASEL</i>
<i>ATRX</i>	<i>FANCC</i>	<i>MITF</i>	<i>RUNX1</i>
<i>AURKA</i>	<i>FANCD2</i>	<i>MLH1</i>	<i>SBDS</i>
<i>AXIN2</i>	<i>FANCE</i>	<i>MPL</i>	<i>SDHA</i>
<i>BAP1</i>	<i>FANCF</i>	<i>MRE11A</i>	<i>SDHAF2</i>
<i>BARD1</i>	<i>FANCG</i>	<i>MSH2</i>	<i>SDHB</i>
<i>BCL2</i>	<i>FANCI</i>	<i>MSH6</i>	<i>SDHC</i>
<i>BCL2L11</i>	<i>FANCL</i>	<i>MUTYH</i>	<i>SDHD</i>
<i>BCL9</i>	<i>FANCM</i>	<i>MYC</i>	<i>SH2D1A</i>
<i>BIRC5</i>	<i>FAS</i>	<i>MYCN</i>	<i>SLX4</i>
<i>BLM</i>	<i>FAT1</i>	<i>NBN</i>	<i>SMAD3</i>
<i>BMPR1A</i>	<i>FBXW7</i>	<i>NF1</i>	<i>SMAD4</i>
<i>BRAF</i>	<i>FGFR1</i>	<i>NF2</i>	<i>SMARCA4</i>
<i>BRCA1</i>	<i>FGFR2</i>	<i>NKX2-1</i>	<i>SMARCB1</i>
<i>BRCA2</i>	<i>FGFR3</i>	<i>NOTCH1</i>	<i>SMC1A</i>
<i>BRIP1</i>	<i>FGFR4</i>	<i>NRAS</i>	<i>SMO</i>
<i>BUB1B</i>	<i>FH</i>	<i>NSD1</i>	<i>SOS1</i>
<i>CARD11</i>	<i>FHIT</i>	<i>NTRK1</i>	<i>SPEN</i>
<i>CASP8</i>	<i>FLCN</i>	<i>PALB2</i>	<i>SPOP</i>
<i>CCND1</i>	<i>GATA2</i>	<i>PAX7</i>	<i>SPRED1</i>
<i>CDH1</i>	<i>GNA11</i>	<i>PDGFB</i>	<i>STAG2</i>
<i>CDK4</i>	<i>GNAQ</i>	<i>PDGFRA</i>	<i>STK11</i>
<i>CDKN1A</i>	<i>GNAS</i>	<i>PDGFRB</i>	<i>SUFU</i>
<i>CDKN1B</i>	<i>H3F3A</i>	<i>PHOX2B</i>	<i>TCF7L2</i>
<i>CDKN2A</i>	<i>HNF1A</i>	<i>PIK3CA</i>	<i>TERT</i>
<i>CDKN2B</i>	<i>HOXB13</i>	<i>PIK3CD</i>	<i>TGFBR2</i>
<i>CDKN2C</i>	<i>HRAS</i>	<i>PIK3R1</i>	<i>TMEM127</i>
<i>CHEK2</i>	<i>IDH1</i>	<i>PIK3R2</i>	<i>TNFAIP3</i>
<i>CREBBP</i>	<i>IDH2</i>	<i>PML</i>	<i>TP53</i>
<i>CTNNA1</i>	<i>IGF2</i>	<i>PMS2</i>	<i>TP63</i>
<i>CTNNB1</i>	<i>IKZF1</i>	<i>POLE</i>	<i>TRIM33</i>
<i>CYLD</i>	<i>INHA</i>	<i>POLH</i>	<i>TSC1</i>
<i>DDB1</i>	<i>INHBA</i>	<i>POT1</i>	<i>TSC2</i>
<i>DDB2</i>	<i>INSR</i>	<i>PPM1D</i>	<i>VHL</i>
<i>DICER1</i>	<i>IRF4</i>	<i>PRKAR1A</i>	<i>WRN</i>
<i>DNMT3A</i>	<i>IRS1</i>	<i>PTCH1</i>	<i>WT1</i>
<i>EGFR</i>	<i>JAK1</i>	<i>PTEN</i>	<i>XPA</i>
<i>EGLN1</i>	<i>JAK2</i>	<i>PTPN11</i>	<i>XPC</i>
<i>EPCAM</i>	<i>JAK3</i>	<i>PTPRD</i>	<i>XRCC2</i>
			<i>XRCC3</i>

Tableau 9 : Liste des 201 gènes par ordre alphabétique

La capture a été modélisée en sommant de manière cumulative pour chaque gène la taille de la capture. Nous avons retenu pour chaque gène la séquence codante (CDS) additionnée pour chaque exon des séquences introniques bordantes +/- 120 paires de bases. La profondeur de lecture a été calculée en divisant la capacité du NextSeq (120 Gb) par la taille cumulative des séquences codantes, pondéré d'un facteur 0,5 correspondant au *on Target* estimé et d'un facteur 0,8 correspondant au Q-score estimé.

$$\sum_{\text{gènes}} = 120 \cdot 10^9 / (\text{CDS} \times N_{\text{exons}} \times 120 \times 2) \times 0,5 \times 0,8$$

Nous avons ensuite modélisé la profondeur en fonction du nombre de gènes retenus et du nombre de patients séquencés par *run* de séquençage. Nous avons ainsi retenu au terme de ce travail bibliographique et de ces compromis une liste de 201 gènes, permettant une profondeur de 2000x pour des runs de 20 patients. Les sondes couvrant l'ensemble des transcrits RefSeq des 201 gènes ont été modélisées à l'aide du logiciel *SureDesign* (Agilent, Santa Clara, CA, USA). Les séquences riches en GC ont été renforcées afin d'améliorer le *on target*. La couverture a été vérifiée *in silico* pour les 201 gènes. La capture représente environ 1 Mb de données de séquençage.

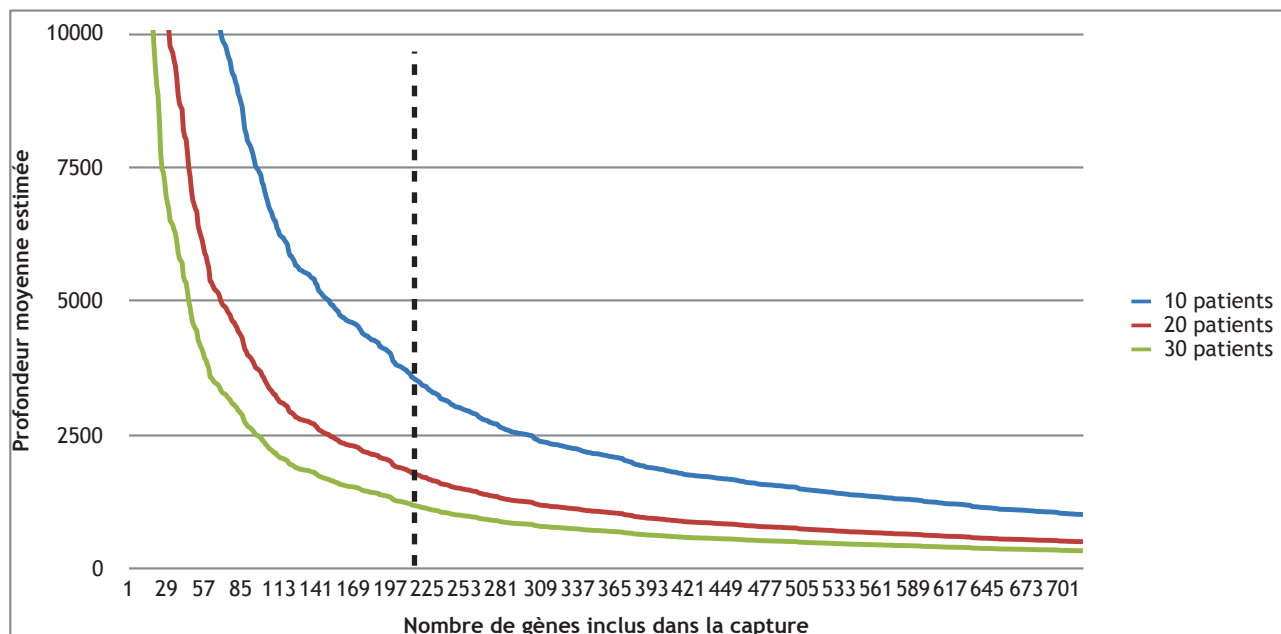


Figure 17 : Modélisation du nombre de gènes inclus dans la capture en fonction des contraintes de profondeur et du nombre de *runs* nécessaires

Préparation des librairies et séquençage

Les librairies d'ADN ont été réalisées selon la technique QXT (*SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing*), à partir d'input de 50 ng d'ADN de qualité dosé par fluorométrie (*TwinckleTM LB970, Berthold, France*). Les séquences d'intérêt ont été capturées à l'aide de la capture 201 gènes. Le séquençage a été effectué par *batch* de 20 patients en plusieurs *runs* de séquençage, sur le séquenceur NextSeq 500, en *paired-end* 2x150pb. Les différentes cohortes de patients (cancer du sein, cancer colorectal, cancer de l'ovaire et cancers pédiatriques) ont été *poolés* lors de ces divers *runs*.

Analyses bio informatiques

Le démultiplexage a été effectué par le logiciel CASAVA version 1.8 d'Illumina. L'alignement et le *base calling* ont été effectués par la suite BWA - Picard - GATK. Un double pipeline a été défini pour le *variant calling* : GATK HaplotypeCaller avec double annotation par Alamut-Batch et Annovar, et réalisation en parallèle d'un second *variant calling* par VarScan2. VarScan2 est à l'origine un algorithme de détection des variants somatiques, détourné ici pour la détection d'évènement en mosaïque [279]. Le logiciel CANOES a également été utilisé pour la détection de CNV [280].

Les variations d'intérêt retenues pour l'analyse correspondent à des altérations associées à un effet biologique important :

1. LOF (*loss-of-function*) : variation non-sens ou affectant un site canonique d'épissage ou *indel frameshift*
2. DSD (*disruptive or strictly damaging*) : incluant les LOF et les variants faux-sens prédits délétères par 3/3 des logiciels de prédiction (SIFT, PolyPhen-2 et MutationTaster)
3. MS (*Missense*) : incluant les variants faux-sens prédits délétères par au moins 2/3 des logiciels de prédiction

Un second filtre permet de retenir uniquement les variations rares associées à une MAF <0,01 dans une population ExaC_All ou de type européenne (ExaC_NFE), population ESP et 1000 Genome. Un troisième filtre correspond à des critères de qualité (Q-Score >0,8 et profondeur >20x) et à une balance allélique comprise entre 20 et 80% afin de réduire les artéfacts tout en permettant l'éventuelle détection d'évènements en mosaïque. Le double pipeline incluant VarScan2 permet néanmoins de rechercher des mosaïques associées à des balances alléliques inférieures à 20% dans une seconde approche plus focalisée sur cette hypothèse.

Les variations identifiées au terme de cette analyse bio informatique ont fait l'objet d'un contrôle sur fichier BAM et d'un séquençage par méthode Sanger.

L'approche statistique visant à rechercher un enrichissement en variants rares dans notre population, comparativement à une population contrôle issue du projet France Exome (FREX), a nécessité un recalibrage afin d'exclure les patients en situation divergente (notamment d'ethnicité), et afin de réduire le biais inhérent aux différences de capture : capture ciblée nous concernant, capture de l'exome concernant le projet France Exome. L'étude de qualité des patients (QC_{patients}) s'est donc appuyé sur les critères suivants :

1. Pourcentage de génotype manquant par individu
2. Ratio transition / transversion
3. Ratio Hétérozygote / Homozygote
4. Nombre de singletons
5. Analyses en composante principale (ACP) portant sur les variants fréquents (MAF>5%) ou rares (MAF<5%)
6. Calcul de l'identity-by-state (IBS) par la méthode de PLINK afin d'écarter tout doublons ou apparentés.

Nous avons également appliqué un filtre sur la qualité des variants (QC_{variants}) sur la base des critères suivants :

1. VQSLOD >99,5% pour les SNV ou >99% pour les *indels*
2. Test d'équilibre d'Hardy-Weinberg excluant les variants lorsqu'en déséquilibre significatif ($p=10^{-7}$) dans la population contrôle
3. Balance allélique incohérente (<90 pour les variants homozygotes ou <25 / >75 pour les variants hétérozygotes)
4. Génotype manquant pour un variant donné pour plus de 5% de l'ensemble de la population
5. Pourcentage de génotypes manquants significatif pour un variant donné entre les patients et les témoins selon un test de Fisher ($p<10^{-7}$).

Nous avons réalisé une estimation des *odds ratio* correspondant à la probabilité chez les patients de porter un variant rare (LOF - DSD ou MS) comparé aux témoins, par un test exact de Fisher. Le seuil de significativité applicable ici est le seuil *exome-wide* par gène de $2,5 \cdot 10^{-6}$ (seuil de 5% corrigé pour une analyse de 20.000 gènes), dès lors que l'on applique les ajustements de Bonferroni. Cette exigence est néanmoins discutable, comme nous l'envisagerons plus après.

Aspects réglementaires

Compte tenu des implications éthiques et de l'exploitation d'une collection biologique à des fins de recherche, cette étude a fait l'objet d'une approbation par le Comité d'Ethique et de Protection des Personnes du Nord-Ouest I et d'une déclaration sur *Clinicaltrial.gov* (NCT02664389). En raison de la cohérence de cette approche avec la finalité initiale du consentement des patients lors de leur prélèvement biologique (identification d'une variation causale dans le contexte de leur pathologie), une information complémentaire aux patients n'a

pas été jugée requise pour leur inclusion dans cette étude dès lors que le consentement initial était archivé et ne comportait pas d'opposition à l'utilisation de la ressource biologique à des fins de recherche. Les correspondants cliniciens en charge de ces patients ont néanmoins été informés de cette étude compte tenu des implications éventuelles sur la prise en charge des patients. En effet, en cas d'identification d'une variation délétère et en présence d'éléments de preuve jugés suffisants pour en établir la causalité, une information du clinicien est nécessaire.

Le protocole rédigé pour cette enregistrement est joint en Annexe.

Résultats

Patients

Nous avons constitué une cohorte rouennaise de 30 patientes affectées par un cancer du sein infiltrant avant 31 ans, après contrôle des données cliniques (caractère infiltrant, négativité des analyses *BRCA1*, *BRCA2* et *TP53*, disponibilité et qualité de l'ADN). La moyenne d'âge était de 26 ans, 28% des patientes présentaient une histoire familiale de cancer au 1^{er} ou 2^{ème} degré, tout cancer confondu.

Les analyses statistiques visant à établir un éventuel enrichissement en variants rares comparativement à la population contrôle FREX se sont appuyées sur une analyse *poolée* associant les cancers du sein précoces séquencés ici par moi-même (n=30) et par Etienne Muller à Caen (n=75) et une analyse *poolant* l'ensemble des cohortes (sein - ovaire - colo-rectal et pédiatrique, (n=307). Les témoins issus de la cohorte FREX sont au nombre de 552. Le contrôle de qualité (QC_{patients}) appliqué à l'ensemble des patients des différentes cohortes et aux témoins, sur les items précédemment développés, révèle :

- Génotype manquant : pas de divergence aberrante
- Ratio transition / transversion (ti/tv) : pas de divergence aberrante
- Ratio Hétérozygote / homozygotes : pas de divergence aberrante
- Nombre de singletons : un nombre élevé pouvant faire évoquer l'hypothèse d'une divergence ethnique par rapport à la population témoin, 19 patients sont exclus de l'analyse statistique dont 3 patientes appartenant à ma cohorte des cancers du sein de moins de 31 ans, en raison d'origines ethniques documentées dans le dossier clinique, africaine ou asiatique. Les analyses en composante principale, effectuées pour l'une sur la base de 127 SNVs fréquents (MAF>5%) et pour l'autre sur 3367 SNVs rares (MAF<5%), confirment ces divergences de population et l'exclusion nécessaire de ces 19 individus. L'analyse statistique finale a porté de ce fait sur 288 patients toutes cohortes confondues.

Variants rares retenus

Le séquençage des 30 patientes a été réalisé sur 4 *runs* de bonne qualité avec un *on target* de 75%, une profondeur de 2200x, un Q-score supérieur à 30 pour plus de 99,9% des

bases. Après application des filtres bio informatiques selon les critères détaillés dans la méthodologie et cure manuelle des artéfacts, nous avons retenu 24 variants rares d'intérêt avec 4 redondances du variant *RAD51C*: c.859A>G, p.T287A, ainsi que deux redondances pour *FANCL* de l'indel c.1114_1115insATTA, p.T372fs (Tableau 7). Ces variants sont portés par 18 patients (60%), et 9 patients comportent plus d'un variant rare.

On remarque avec intérêt que sur ces 24 variants rares, 16 concernent des gènes impliqués dans la réparation de l'ADN, suggérant un enrichissement particulier de cette voie biologique dans notre population. Ces 16 gènes impliquent tant la voie de la réparation homologue (*RAD51C*, *CHEK2*, *RINT1*, *WRN*, *FANCI*, *FANCL* et *FANCM*), que les voies de la réparation par excision nucléotidique (NER) avec *XPC*, *SLX4*, *POLH*, *ERCC2*, *ERCC3* et *DDB2*, la voie de la réparation par excision de base (BER) avec *MUTYH* et la voie de réparation des mésappariements (*mismatch repair*) avec *MLH1*, et comme attendu *TP53*. Si la conception de la capture est biaisée par l'intérêt majeur porté à ces différentes voies en raison de leur implication tant somatique que constitutionnelle dans la survenue des cancers (48 gènes sur 201), nous identifions néanmoins sur cette simple constatation un enrichissement significatif ($p < 0,00001$, Test de Fisher).

Nous avons procédé à une analyse fine de ces 24 variants rares, basées sur les critères de qualité après analyse des BAM, vérification des fréquences en population générale dans les bases de données, vérification des prédictions bio informatiques et des données de conservation sur Alamut, interrogation des bases de données générales type ClinVar ou gène-spécifiques et interprétation au regard de la littérature.

Patient	Gène	NM	c.	p.	Classe
#1	<i>XPC</i>	NM_004628	c.1643_1644del	p.(Val548X)	LOF
#2	<i>RAD51C</i>	NM_058216	c.859A>G	p.(Thr287Ala)	MS
#3	<i>SLX4</i>	NM_032444	c.3374A>G	p.(Asp1125Gly)	MS
	<i>XPC</i>	NM_004628	c.1933C>G	p.(Pro645Ala)	MS
	<i>RET</i>	NM_020630	c.874G>A	p.(Val292Met)	MS
#4	<i>MUTYH</i>	NM_001048171	c.1145G>A	p.(Gly382Asp)	MS
	<i>POLH</i>	NM_001291970	c.40A>G	p.(Met14Val)	MS
#5	<i>ERCC3</i>	NM_000122	c.371A>T	p.(Tyr124Phe)	MS
#6	<i>CHEK2</i>	NM_145862	c.1340C>T	p.(Thr447Met)	MS
#7	-		-	-	-
#8	-		-	-	-
#9	-		-	-	-
#10	<i>MLH1</i>	NM_000249	c.347C>A	p.(Thr116Lys)	MS
#11	<i>ARID1A</i>	NM_006015	c.6706C>T	p.(Arg2236Cys)	MS

Patient	Gène	NM	c.	p.	Classe
#12	-		-	-	-
#13	RAD51C	NM_058216	c.859A>G	p.(Thr287Ala)	MS
#14	ERCC3	NM_000122	c.1048G>A	p.(Gly350Ser)	MS
	APC	NM_001127511	c.2447C>T	p.(Ser816Phe)	MS
#15	RINT1	NM_021930	c.1985T>C	p.(Leu662Ser)	MS
	TSC1	NM_001162427	c.3184C>T	p.(Arg1062Trp)	MS
	FANCL	NM_001114636	c.1114_1115insATT	p.(Thr372X)	LOF
#16	RAD51C	NM_058216	c.859A>G	p.(Thr287Ala)	MS
#17	ERCC2	NM_001130867	c.1105C>T	p.(Arg369Cys)	MS
	RAD51C	NM_058216	c.859A>G	p.(Thr287Ala)	MS
#18	-		-	-	-
#19	-		-	-	-
#20	-		-	-	-
#21	FANCL	NM_001114636	c.1114_1115insATT	p.(Thr372X)	LOF
	DDB2	NM_001300734:	c.550G>C	p.(Asp184His)	MS
#22	MET	:NM_000245	c.2975C>T	p.(Thr992Ile)	MS
	FANCI	NM_001113378	c.1813C>T	p.(Leu605Phe)	MS
#23	KIT	NM_000222	c.2279A>T	p.(Asp760Val)	MS
#24	WRN	NM_000553	c.3785C>G	p.(Thr1262Arg)	MS
	TP53	NM_001126115	c.1024C>T	p.(Arg342X)	LOF
#25	-		-	-	-
#26	FANCM	NM_020937	c.497C>A	p.(Ala166Asp)	MS
	CHEK2	NM_145862	c.1036C>T	p.(Arg346Cys)	MS
#27	-		-	-	-
#28	-		-	-	-
#29	-		-	-	-
#30	-		-	-	-

Tableau 10 : variants rares identifiés dans la cohorte de cancers du sein précoce

La **patiente #1** est une jeune patiente de 30 ans, d'origine nord-africaine, affectée d'un cancer du sein dont le phénotype n'a pas été communiqué, présentant des antécédents familiaux au premier degré. La variation *XPC* détectée (c.1643_1644del, p.(Val548X)) est une délétion perte de fonction par décalage du cadre de lecture. La variation est lue à une profondeur de 2470x, portée par 75% des *reads*. Cette variation est rapportée dans la population ExaC_all comme rare, avec une fréquence de $8,3 \cdot 10^{-6}$. De manière intéressante, cette variation est décrite à l'état récessif comme une variation délétère fondatrice tunisienne du *xeroderma pigmentosum* de type 3 (XP3), avec l'identifiant ClinVar RCV000000286.3. Cette altération tronque le domaine C-terminal de la protéine *XPC* (résidus 492-940), lequel interagit normalement avec l'ADN et les protéines RAD23B, CETN2 et TFIIH. Cette variation à l'état récessif induit de ce fait un défaut du système de réparation *nucleotide-excision repair* (NER). Le *xeroderma pigmentosum* (OMIM #

278720) se manifeste par une sensibilité accrue à la génotoxicité des rayons ultra-violet, à l'origine de cancers cutanés précoces et récidivants. Compte tenu de la pathogénicité des altérations de *XPC* à l'état récessif, nous avons recherché infructueusement sur l'ensemble du gène d'autres altérations, polymorphismes ou variations introniques profondes. L'origine ethnique de cette patiente et la fréquence du variant V548fs dans la population tunisienne [281] laisse supposer que cette patiente soit hétérozygote saine. Néanmoins, on note dans la littérature des études d'association entre des variants d'*XPC*, tant dans un modèle récessif que dominant, et la survenue de cancers non cutanés (poumon, vessie et colorectal) [282]. Par ailleurs, il est évoqué un rôle protecteur d'*XPC* face au stress oxydatif induit par les œstrogènes, faisant proposer un modèle de carcinogenèse mammaire que supporte la diminution d'expression observée dans le cancer du sein [283]. Une contribution de cette variation délétère, à l'état hétérozygote, à la survenue précoce d'un cancer du sein ne peut donc être exclue.

La **patiente #2** présente également un cancer du sein à 30 ans, sans histoire familiale de cancer. La variation de *RAD51C* c.859A>G, p.(Thr287Ala) est détectée sur 44% des *reads* à une profondeur de 3173x. La fréquence ExaC_all rapportée est néanmoins fréquente (0,55%) et la variation est rapportée dans ClinVar comme bénigne (RCP000130980.3). Cette variation est également détectée chez les patientes #13, #16 et #17. Les prédictions bio informatiques sont discordantes, avec notamment une tolérabilité prédite par *SIFT*. Cependant, cette variation est rapportée comme associée (de manière non significative) avec la survenue d'un cancer du sein (OR=5,5) [284]. Néanmoins, comme exposé en introduction, la contribution des altérations des paralogues de *RAD51* au cancer du sein est débattue [137]. En outre, ce variant est rapporté en tant que polymorphisme dans la population danoise [285]. La pathogénicité de ce variant n'est donc pas retenue au terme de ces différentes observations.

La **patiente #3** présente 3 altérations faux-sens d'intérêt dans le contexte d'un cancer du sein à 24 ans, de présentation sporadique. On notera l'origine asiatique pour cette patiente, conduisant à son exclusion dans l'approche statistique comparant notre population à la cohorte FREX. L'altération faux-sens de *SLX4* c.3374A>G, p.(Asp1125Gly) est présente à l'état hétérozygote (AB=48%) et détectée à une profondeur de 3597x. Ce variant est rare dans les bases de données avec une fréquence ExaC_all de $2,5 \cdot 10^{-5}$. On notera cependant une fréquence 10 fois plus importante dans la population asiatique ($3 \cdot 10^{-4}$). Ce variant est prédit délétère par les 3 logiciels de prédiction, et affecte un acide aminé conservé jusqu'au *xenope*. Le gène *SLX4* (*FANCP*) est impliqué dans la maladie de Fanconi à l'état homozygote (OMIM #613951). L'altération est cependant ici strictement hétérozygote, sans altération retrouvée en *trans* après vérification des BAM. Des altérations de *SLX4* ont cependant été retrouvées dans des formes héréditaires de cancer du sein (non-sens=1 ; faux-sens prédites délétères=22), avec un test de complémentation en faveur de la pathogénicité uniquement pour l'altération non-sens c.2469G>A, p.(Trp823X) [179]. Cette variation est donc peu probablement pathogène dans ce contexte. La seconde variation est à nouveau une altération faux-sens du gène *XPC* : c.1933C>G, p.(Pro645Ala). La balance allélique est en faveur d'un statut hétérozygote (AB=42%), avec une

profondeur de lecture de 2846x. La fréquence ExaC_all est $2,5 \cdot 10^{-5}$, les prédictions sont unanimement en faveur d'une pathogénicité. Les mêmes arguments que pour la patiente #1 en faveur d'une implication dans la carcinogenèse peuvent être évoqués chez cette patiente. Enfin, une altération du gène RET: c.874G>A, p.(Val292Met) est retrouvée sur 50% des *reads* à une profondeur de 1813x. La fréquence en population totale est faible (ExaC_all = $6 \cdot 10 \cdot 10^{-4}$) mais plus importante en population asiatique ($6 \cdot 10^{-3}$). Les assertions ClinVar sont discordantes (variant neutre, de signification inconnue ou pathogène dans la néoplasie endocrinienne multiple type 2). Les prédictions sont en faveur d'un effet délétère, la substitution affectant un acide aminé hautement conservé jusqu'au *xenope* et intervenant au niveau du domaine tyrosine kinase. On ne retrouve pas dans l'histoire familiale d'argument pour une néoplasie endocrinienne multiple, et l'implication de cette variation soulèverait l'hypothèse d'une extension phénotypique, non retrouvée dans la littérature dans la description des familles affectées par la NEM2A. Néanmoins, une variation délétère du gène *RET* (sur 282 individus) a été rapportée dans une analyse par panel d'une population de femmes israéliennes atteintes de cancer du sein héréditaire [286]. L'ensemble de ces éléments ne permet donc pas de retenir cette variation de *RET* comme d'intérêt.

La **patiente #4** présente deux altérations faux-sens : *MUTYH*:c.1145G>A, p.(Gly382Asp) et *POLH*:c.40A>G, p.(Met14Val), dans un contexte de cancer du sein sporadique à 29 ans sans description de la situation familiale. L'altération de *MUTYH* est retrouvée hétérozygote (AB=49%) à une profondeur de lecture de 4703x. La fréquence ExaC_all est de $2,8 \cdot 10^{-3}$. Les prédictions sont uniformément en faveur de la pathogénicité et la position concernée par la substitution est hautement conservée. Deux entrées ClinVar sont référencées en faveur d'une syndrome de prédisposition héréditaire au cancer (RCV000115748.5) ou d'une polypose familiale atténuée type 2 (FAP type 2) (RCV000005614.9). Néanmoins, dans la polypose atténuée, les altérations sont nécessairement bi-alléliques. L'analyse sur le fichier BAM de l'ensemble du gène *MUTYH* ne révèle pas d'autre altération. Le gène *MUTYH* est impliqué dans le système de réparation BER, associé à une signature tumorale marquée par des transversions C:G>A:T. Néanmoins, les variations mono-alléliques de *MUTYH* sont retrouvées non associées au risque de cancer du sein dans une vaste étude (30.000 cas et 12.000 témoins) [287], mettant un terme à plusieurs études évoquant un sur-risque pour les porteurs mono-alléliques de variants pathogènes de *MUTYH* [288, 289]. L'altération de *POLH* est également hétérozygote (AB=47%), avec une profondeur de lecture de 2921x. La fréquence ExaC_all est de $2 \cdot 10^{-4}$. Les prédictions sont en faveur de la pathogénicité pour les 3 algorithmes avec une forte conservation tant nucléotidique que protéique. Le gène *POLH* (ou *XPV*) est impliqué dans une forme variante atténuée du *xeroderma pigmentosum* (OMIM #603968). Des altérations faux-sens de *POLH* ont été rapportées dans l'étude GENESIS [216], mais avec des *odds ratio* inférieur à 1, en faveur d'un effet protecteur vis-à-vis de la survenue d'un cancer du sein.

La **patiente #5**, affectée d'un cancer du sein sporadique à 29 ans, présente une variation faux-sens du gène *ERCC3*: c.371A>T, p.(Tyr124Phe). Cette variation est présente à l'état

hétérozygote (AB=50%) à une profondeur de 3656x. Aucune fréquence ExaC n'est rapportée pour ce variant, qui affecte des positions nucléotidiques et en acide aminé extrêmement conservées, intéressant le domaine *helicase*. Les prédictions sont en faveur d'un effet délétère pour *SIFT* et *MutationTaster*, *possibly damaging* uniquement pour *Polyphen-2*. *ERCC3* appartient de nouveau au système NER, et les altérations de ce gène sont à l'origine à l'état récessif d'une variante du *Xeroderma pigmentosum* (groupe B) incluant la trichothiodystrophie ou le syndrome de Cockayne (OMIM: 616390 et 610651). Nous n'avons pas identifié d'autres altérations d'*ERCC3* sur le fichier BAM pouvant être à l'origine d'une hétérozygotie composite. La patiente ne présente pas par ailleurs d'argument en faveur d'une forme de *Xeroderma pigmentosum*. La contribution d'*ERCC3* à la survenue de ce cancer du sein précoce est donc incertaine, même si une association modérée a été retrouvée entre une variation tronquante d'*ERCC3* et la survenue de cancers du sein dans une population ashkénaze [290].

La **patiente #6**, affectée par un cancer du sein sporadique à 22 ans, présente le variant *CHEK2*: c.1340C>T, p.(Thr447Met) à l'état hétérozygote (AB=51%) sur une profondeur de lecture de 1370x. La fréquence ExaC de ce variant est de $5,4 \cdot 10^{-4}$, soit un variant à la limite de la définition d'un variant rare (10^{-3}). Ce variant est rapporté de multiples fois dans ClinVar comme *uncertain significance/ likely pathogenic*. Compte tenu des éléments d'interprétation de ce variant rapportés dans la littérature et de sa fréquence élevée, nous ne considérons pas ce variant comme relevant dans la causalité d'un cancer du sein précoce.

La **patiente #10**, affectée d'un cancer du sein sporadique triple négatif à 24 ans, présente l'altération faux-sens de *MLH1*: c.347C>A, p.(Thr116Lys). Cette variation est retrouvée à l'état hétérozygote (AB=46%) à une profondeur de lecture de 1522x. Elle affecte une position conservée et le domaine protéique de liaison à l'ADN et le domaine *ATPase*. Les prédictions sont discordantes (2/3) concernant la prédiction de pathogénicité. Bien que non rapportée dans la base ExaC, cette variation est rapportée comme de signification indéterminée dans ClinVar et dans la base InSiGHT, et sans effet sur l'épissage retrouvé dans une approche *minigene* [291]. Nous ne retenons donc pas cette variation comme possiblement causale.

La **patiente #11** affectée d'un cancer du sein avec récepteurs hormonaux positifs à 30 ans, dans une présentation sporadique, présente le variant faux-sens d'*ARID1A*: c.6706C>T, p.(Arg2236Cys). Ce variant est hétérozygote (AB=47%), retrouvé à une profondeur de 3164x. Les prédictions sont en faveur d'un effet délétère (3/3), et la substitution concerne une position hautement conservée au niveau du dernier exon. La fréquence ExaC rapportée est de $2,4 \cdot 10^{-5}$, et ce variant est rapporté dans dbSNP avec le rs763691986. *ARID1A* participe au complexe *SWI/SNF*, et est retrouvé altéré au niveau somatique dans un nombre conséquent de types tumoraux, notamment dans les cancers de l'ovaire à cellules claires et endométrioides. Des données d'expression dans le cancer du sein supportent l'éventuelle implication d'*ARID1A* dans la carcinogenèse mammaire [292, 293]. Les altérations tronquantes ou délétions hétérozygotes d'*ARID1A* sont à l'origine du syndrome de Coffin-Siris type 2 (OMIM: 603024) incluant un retard

mental et un syndrome malformatif, non observé chez ce cas. L'ensemble de ces observations conduit à considérer cette altération comme possiblement causale dans la pathologie mammaire.

La **patiente #14**, affectée d'un cancer du sein sporadique triple négatif à 23 ans, présente 2 variants : *ERCC3*:c.1048G>A, p.(Gly350Ser) et *APC*: c.2447C>T, p.(Ser816Phe). Le variant d'*ERCC3* est présent à l'état hétérozygote (AB=46%) à une profondeur de lecture de 2660x. Il est prédit délétère par les 3 algorithmes, la substitution affecte une position conservée. La fréquence ExaC_all est $1,6 \cdot 10^{-5}$. L'argumentaire développé pour la patiente #5 peut être également proposé ici, avec l'absence d'hétérozygotie composite après analyse des BAM pour ce gène à l'origine du *Xeroderma pigmentosum* groupe B lorsqu'affecté à l'état homozygote, et la modeste association retrouvée entre des variations tronquantes d'*ERCC3* et la survenue de cancers du sein dans la population ashkénaze [290]. La variation d'*APC* est également hétérozygote (AB=47%), retrouvée à la profondeur de lecture de 2378x. Les prédictions (3/3) sont en faveur d'un effet délétère, avec une substitution affectant un acide aminé moyennement conservé au niveau du dernier exon. Cette variation n'est pas associée à une fréquence dans ExaC et n'est pas rapportée dans ClinVar ni dans *UMD-APC*, notamment comme à l'origine de la polypose adénomateuse familiale. La contribution de ces deux variants rares pour cette patiente n'est donc pas retenue.

La **patiente #15** est une jeune femme de 20 ans affectée d'un cancer du sein sporadique récepteurs hormonaux positifs, HER2 négatif. Nous avons identifié 3 variants d'intérêt chez cette patiente sur les gènes *RINT1*, *TSC1* et *FANCL*. La variation *RINT1*: c.1985T>C, p.(Leu662Ser) est présente à l'état hétérozygote (AB=47%) et une profondeur de lecture de 1354x. La substitution, bien qu'affectant une position modérément conservée (poulet), intervient dans le domaine protéique RINT1/TIP1. Les prédictions (3/3) sont en faveur d'un effet délétère. La fréquence ExaC_all est $1,2 \cdot 10^{-4}$. La contribution des variations constitutionnelles de *RINT1* dans la survenue de cancers du sein héréditaire avait été soulevée par la publication de Park *et al* [173], puis infirmée par la large série de Li *et al* [174]. Ce variant n'est donc pas retenu comme étant d'intérêt. La variation *TSC1*: c.3184C>T, p.(Arg1062Trp) est également hétérozygote, détectée à une profondeur de lecture de 3357x. Les prédictions (3/3) sont en faveur d'un effet délétère, avec une substitution affectant une position hautement conservée. La fréquence ExaC est de $1 \cdot 10^{-4}$, et cette variation est décrite à l'état somatique dans la base Cosmic. Néanmoins l'insertion RCV000189827.1 correspondant à ce variant dans ClinVar est en faveur d'une classe 2 (*likely benign*). Par ailleurs, il n'est pas retrouvé dans la population affectée par la sclérose tubéreuse de Bourneville (OMIM: 191100), syndrome lié à des variations autosomiques dominantes de *TSC1* ou *TSC2*, de risque accru de pathologie maligne excepté pour le cancer du rein [294]. On notera cependant que les âges d'apparition des tumeurs dans cette population sont davantage précoces (40,5 vs. 66 ans). *FANCL* est affecté par une *indel* non en phase : c.1114_1115insATTA, p.(Thr372X). Cette *indel* affecte le dernier exon de *FANCL*, relativement mal couvert dans notre capture avec une profondeur de 843x. Il s'agit d'un évènement non rare, avec une fréquence ExaC de 0,29% en population générale. Bien que décrit chez des patients évocateurs d'anémie de Fanconi groupe L (SCV000247355), cette *indel* est aussi interprétée comme de signification

inconnue (SCV000290416.1). A noter cependant que cette *indel* est également présente chez la patiente #21, soit une récurrence notable au sein de notre série.

La **patiente #17**, affectée à 25 ans d'un cancer du sein, présente le variant *ERCC2*: c.1105C>T, p.(Arg369Cys) à état hétérozygote (AB=47%) et une profondeur 1649x. Les prédictions bio informatiques sont en faveur d'un effet délétère, la substitution affectant le domaine *helicase* et une position très conservée. Aucune fréquence ExaC n'est rapportée pour ce variant, avec un z-score pour les variations faux-sens à 0,44. A l'instar d'*ERCC3*, *SLX4* et *XPC*, *ERCC2* appartient au système de réparation NER, et est impliqué à l'état récessif dans le *Xeroderma pigmentosum*, groupe D (OMIM: 278730). Nous n'avons pas identifié d'autre variation sur *ERCC2* chez cette patiente pouvant aboutir à un état récessif par hétérozygotie composite. *ERCC2* a fait l'objet d'une approche gène-candidat dans une population d'Europe Centrale évocatrice d'une prédisposition héréditaire au cancer du sein et de l'ovaire [295]. Le séquençage de 1345 cas index a ainsi révélé 3 variations non-sens, 2 variations d'épissage et 20 variations faux-sens rares, n'incluant pas notre variation identifiée. Des analyses fonctionnelles de ces variants sont conduites avec un test de complémentation. Quatre variants sur 11 testés révèlent ainsi un déficit d'activité de la voie NER. La majorité de ces variants apparaissent cependant dans cette population comme des variations privées, et la co-ségrégation des variants avec le phénotype n'est pas démontrée. Cette étude soulève ainsi la question d'une possible implication d'*ERCC2* dans la survenue de cancers héréditaires, tout en soulignant la difficile démonstration de cette causalité.

La **patiente #21**, affectée à 29 ans, présente le variant *DDB2*: c.550G>C, p.(Asp184His) à l'état hétérozygote (AB=47%) et à une profondeur de lecture de 2285x. Les 3 algorithmes de prédictions sont en faveur d'un effet délétère, avec une position très conservée jusqu'à la levure. Cette variation n'est pas rapportée dans ExaC, avec un z-score élevé pour les variations faux-sens. A nouveau, il s'agit ici d'une variation affectant un gène impliqué dans le système NER, dont les altérations à l'état récessif sont à l'origine du *Xeroderma pigmentosum* groupe E (OMIM: 278740). Chez cette patiente également, la vérification de l'ensemble du gène n'identifie pas de second évènement en faveur d'une hétérozygotie composite. Une publication de 2008 soulevait l'hypothèse d'une contribution de *DDB2* en tant qu'oncogène dans le cancer du sein récepteurs hormonaux positif [296]. On notera cependant que la patiente #21 est décrite comme de phénotype RH négatif. Par ailleurs, aucune variation de *DDB2* n'est identifiée dans une cohorte russe de patientes affectées de cancers du sein héréditaire [297]. La contribution de ce variant est donc jugée non conclusive.

La **patiente #22**, affectée à 28 ans, présente deux variants : *MET*: c.2975C>T, p.(Thr992Ile) et *FANCI*: c.1813C>T, p.(Leu605Phe). Le variant de *MET* est associé à une fréquence ExaC de 0,79% et rapporté comme de signification inconnu dans ClinVar. Le variant de *FANCI* est également rapporté avec une fréquence ExaC non rare (0,62%) et présente trois assertions ClinVar en faveur de la bénignité. Ces variants ne sont donc pas retenus comme pertinents.

La **patiente #23**, affectée à 26 ans avec antécédent de cancer du sein au 1^{er} degré, présente le variant *KIT*: c.2279A>T, p.(Asp760Val). Ce variant est observé à l'état hétérozygote (AB=46%) et une profondeur de lecture de 815x. Les prédictions (3/3) sont en faveur d'un effet délétère, la substitution intervenant au niveau du site catalytique. La fréquence ExaC est de 8.10^{-6} et le variant est référencé dans dbSNP par le rs747847018. Néanmoins ce variant est rapporté dans ClinVar comme de signification incertaine (SCV000550117.4), et n'est rapporté qu'une fois dans COSMIC dans une leucémie lymphoïde chronique (COSM4383743). Nous ne retenons donc pas ce variant comme causal dans cette présentation.

La **patiente #24** (ou F332) présente un cancer du sein bilatéral à 27 et 34 ans, avec un antécédent familial au 2^{ème} degré de cancer du sein chez sa grand-mère maternelle. Deux variants sont identifiés : *WRN*: c.3785C>G, p.(Thr1262Arg) et *TP53*: c.1024C>T, p.(Arg342X). La variation de *WRN* est présente à l'état hétérozygote (AB=51%) à une profondeur de lecture de 1050x. Les prédictions (3/3) sont en faveur d'un effet délétère, mais la fréquence ExaC_all est de 0,27% et plusieurs références ClinVar (RCV000122293.1 et RCV000231281.1) ne sont pas en faveur d'un effet délétère. Ce variant n'est donc pas retenu comme causal.

La variation de *TP53* de type non-sens est détectée avec une balance allélique de 17%, sur une profondeur de 4127x, avec une répartition harmonieuse sur les *reads* sens et anti-sens (353+/346-). Cette balance allélique indique une variation présente à l'état de mosaïque : **c.[1024C=>T], p.[R342=/(R342X)]**. La variation est par ailleurs retrouvée sur les séquences Sanger (figure 12). Cette altération trouquante intervient au niveau de l'exon 10, et affecte ainsi les domaines de tétramérisation et la séquence NES. Cette variation est rapportée dans ClinVar comme pathogène (RCV000161074.3 et RCV000213069.1), ainsi que dans la base de l'*IARC* *TP53* avec 21 références à l'état germlinal. Cette variation a en particulier été rapportée dans des cancers du sein précoce (n=8), des corticosurrénalomes (n=3), tumeurs cérébrales (n=3) et sarcomes (n=2). Elle est également rapportée comme un *hotspot* mutationnel au niveau somatique. Cette variation délétère ne présente pas d'effet trans-dominant négatif dans le test fonctionnel à la levure [298, 299]. La présentation phénotypique avec cancer du sein bilatéral et précoce est en adéquation avec la présence d'une variation délétère de *TP53*. Le contexte sporadique de cette patiente, avec mère indemne de cancer et probable phénocopie de la grand-mère maternelle affectée tardivement d'un cancer du sein, est compatible avec l'identification en mosaïque de cette variation. Les cliniciens en charge de cette patiente ont été informés de cette identification, notamment pour l'optimisation de la surveillance de cette patiente et l'adaptation du conseil familial. La patiente n'a pas souhaité faire l'objet de nouveaux prélèvements confirmatoires, notamment de prélèvements urinaires, cutanés ou salivaires pouvant permettre la confirmation d'une atteinte en mosaïque. **Nous retenons donc cette variation comme délétère, en situation de mosaïque probable** compte tenu de la balance allélique à 17%. Cette problématique des variations de *TP53* en mosaïque fait l'objet de la 3^{ème} partie de notre Thèse, compte tenu de son impact clinique majeur.

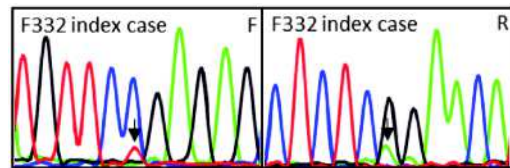


Figure 18 : Visualisation sur ALAMUT et électrophorégramme SANGER de la variation TP53: c.[1024C=>T], p.[R342=/(R342X)]

La **patiente #26**, affectée à l'âge de 26 ans par un cancer du sein récepteurs hormonaux positifs, HER2 négatif, dans une présentation sporadique, présente deux variants faux-sens. La variation *FANCM*: c.497C>A, p.(Ala166Asp) est hétérozygote (AB=49%), détectée à une profondeur de lecture de 1431x. Les prédictions de pathogénicité sont discordantes (2/3), bien que la variation affecte les domaines *DNA/RNA helicase* et *DEAD-like helicase*, et que la substitution intervienne sur une position très conservée (drosophile). Cette variation n'est pas rapportée dans ExaC ni dans ClinVar. Cependant les variations p.(Ala166Trp) ou p.(Ala166Ala) présentent des fréquences ExaC de $1,7 \cdot 10^{-5}$, et le z-score de *FANCM* est négatif (-1,3) indiquant une tolérance aux variations faux-sens. *FANCM* est impliqué dans le complexe de Fanconi, et les variations bi-alléliques sont débattues comme pouvant être à l'origine de l'anémie de Fanconi du groupe M [300]. Son implication est en revanche reconnue dans l'infertilité masculine monogénique avec azoospermie ou oligospermie. Nous n'avons pas retrouvé chez la patiente #26 de second évènement sur le gène *FANCM*. Les variations hétérozygotes faux-sens ou non-sens de *FANCM* ont été recherchées dans des cohortes de cancers du sein masculin [301], ou de cancers du sein triple négatif [302, 303]. Une association significative a été identifiée entre la présence de la variation non-sens c.5101C>T (p.Q1701X) de *FANCM* et la survenue d'un cancer du sein triple négatif dans la population finlandaise. Considérant l'effet de cette variation faux-sens, le z-score de *FANCM* et une littérature controversée quant à l'implication de *FANCM* dans le cancer du sein héréditaire, nous ne retenons donc pas ce variant comme possiblement causal.

La seconde variation concerne le gène *CHEK2* : *CHEK2*: c.1036C>T, p.(Arg346Cys). Cette variation est hétérozygote (AB=48%), détectée à une profondeur de 1438x. Les prédictions de pathogénicité sont en faveur d'un effet délétère (3/3), et la distance de Grantham induite par cette substitution est élevée (180). Le résidu Arg346 est conservé jusqu'à la levure, et participe à de nombreux domaines fonctionnels de *CHEK2*. La fréquence ExaC de cette variation est de 4.10^{-5} , mais cette variation est rapportée comme de signification inconnue à 4 reprises dans ClinVar. Cette variation est également référencée dans dbSNP par le rs201206424. Sur la base des arguments développés pour la patiente #6, l'implication de cette variation de *CHEK2* dans la survenue d'un cancer du sein précoce n'est donc pas retenue.

Nous n'avons pas détecté, après vérification visuelle des BAM et élimination des artéfacts, de réarrangements de grande taille ou CNV par le logiciel CANOES. De même, hormis la variation en mosaïque de *TP53* chez la patiente #24, nous n'avons pas détecté par le logiciel VarScan2 d'autres variations de balance allélique inférieure à 20% permettant de suspecter d'autres événements en mosaïque.

Recherche d'association et d'enrichissements en variants rares spécifiques

Cette approche vise à identifier dans notre population un éventuel enrichissement en variants rares parmi l'ensemble des 201 gènes capturés ou dans certains gènes impliqués dans des voies biologiques d'intérêt, en comparaison avec une population contrôle issue des données du *French Exome Project* (FREX).

La filtration des patients sur les critères de qualité, détaillée en amont, a permis de retenir 228 individus patients toutes cohortes confondues et 552 témoins FREX. La filtration des variants sur les critères de qualité a drastiquement conduit à ne retenir que 53% des variants rares détectés de type LOF, 73% des variants rares lorsqu'on considère les variants rares LOF + DSD. Ce processus de filtration a été comparé à une filtration manuelle pour une liste de 16 artéfacts, avec une excellente corrélation.

Sur l'ensemble des gènes, les *odds ratios* apparaissent en faveur d'une association entre variants rares et survenue d'un cancer précoce, sans toutefois atteindre la significativité requise ($p < 2,5.10^{-6}$) compte tenu des corrections de Bonferroni. Pour exemple, l'association entre variants LOF affectant les gènes de la réparation de l'ADN dans l'ensemble des cohortes est appréciée par un *odds ratio* de 2,45 avec une p-valeur de 6.10^{-2} , et ceux même après restriction aux patients adultes. Compte tenu de l'observation d'un enrichissement dans les gènes de la réparation, nous avons distingué différentes listes de gènes impliqués dans la réparation de l'ADN: Recombinaison Homologue (HR), *Non-Homologous End-Joining* (NHEJ), Interstrand Cross-Linking (ICL), Mismatch Repair (MMR), Nucleotide-Excision Repair (NER) et Base-Excision Repair (BER) et une liste agrégeant l'ensemble de ces gènes. L'association la plus forte (OR = **2,96**, IC95%(1,18-7,00), $p = 1,5.10^{-2}$) est obtenue pour les patients atteints de cancers du sein ou de l'ovaire précoces avec l'ensemble des gènes de la réparation, pour les variants LOF.

Type Variant	Cohorte	Liste de gène	N _{porteurs} / patients	Total patients	N _{porteurs} / contrôles	Total contrôles	OR	IC _{inf}	IC _{sup}	P-valeur
LOF	Adult	DNArepairGene	15	186	20	552	2,33	1,08	4,91	0,02599
LOF	ALL	DNArepairGene	22	288	18	552	2,45	1,23	4,94	0,00613
LOF	Colon	DNArepairGene	5	76	20	552	1,87	0,53	5,35	0,21107
LOF	PED	DNArepairGene	4	87	13	552	2,00	0,46	6,66	0,27048
LOF	SO	DNArepairGene	10	110	18	552	2,96	1,18	7,00	0,01545
DSD	SO	DNArepairGene	45	110	210	552	1,13	0,72	1,74	0,59257
LOF	SO	HR	1	110	5	552	1,00	0,02	9,10	1,00000
LOF	SO	ICL	3	110	5	552	3,06	0,47	15,99	0,13271
LOF	SO	MMR	0	110	0	552	0,00	0,00	Inf	1,00000
LOF	SO	NER	1	110	4	552	1,26	0,03	12,86	1,00000
LOF	Adult	notDNArepairGene	14	186	29	552	1,47	0,70	2,95	0,27748
LOF	ALL	notDNArepairGene	23	288	30	552	1,51	0,82	2,75	0,17795
LOF	Colon	notDNArepairGene	5	76	29	552	1,27	0,37	3,47	0,59007
LOF	PED	notDNArepairGene	7	87	26	552	1,77	0,63	4,36	0,19320
LOF	SO	notDNArepairGene	9	110	27	552	1,73	0,69	3,93	0,16879

Tableau 11 : Associations entre variants rares et survenue d'un cancer précoce.

LOF : variants perte de fonction - DSD : variants LOF + variants faux-sens avec prédiction de pathogénicité prédite par 3 algorithmes - ALL : toutes cohortes confondues - PED : cohorte pédiatrique - SO : cohorte sein + ovaires - DNArepairGene : liste de gènes impliqués dans la réparation de l'ADN - HR : *homologous recombination* - ICL : *Interstrand Cross-Linking* - MMR : *mismatch repair* - NER : *nucleotide-excision repair* - notDNArepairGene : gènes non impliqués dans la réparation de l'ADN - OR = odds ratio

Cette association est en revanche abolie (OR=1,12, IC95%(0,72-1,74) lorsqu'on intègre les variants faux-sens prédits délétères (LOH + DSD) (Tableau 11). Par ailleurs, ces associations sont non significatives, bien en-deçà du seuil de $2,5 \cdot 10^{-6}$, la meilleure p-valeur étant de $6 \cdot 10^{-3}$ pour les variants LOF de la liste agrégée des gènes de la réparation et pour l'ensemble des cohortes. L'hypothèse d'un enrichissement en variants rares chez les patients, parmi les gènes impliqués dans la réparation de l'ADN, en comparaison aux autres gènes de la capture et chez les contrôles, revient au calcul du ratio entre les *odds ratio* « *DNA repair* » et « *Not_DNA repair* ». Ce rapport est de 1,54, ne permettant donc pas d'établir un réel enrichissement.

Discussion

Nous avons basé nos travaux, dans cette approche par capture ciblée, sur deux éléments : une population de **cancers du sein d'extrême précocité est évocatrice d'un déterminisme génétique** indépendamment de l'histoire familiale ou du phénotype tumoral, et l'élargissement d'un **panel de gènes enrichi en gènes impliqués au niveau somatique pourrait permettre la détection de variants rares délétères** au-delà des gènes canoniques de la prédisposition héréditaire au cancer. Nous constatons que la vérification de ces deux hypothèses est plus qu'incertaine, avec 40% des patients sans variant rare prédit délétère et une causalité dans le déterminisme d'un cancer précoce demeurant très putative pour les variants rares perte de fonction ou prédits délétères pour les 60% patients identifiés porteurs. Seule la variation délétère en mosaïque *TP53*: c.[1024C=>T], p.[R342=/(R342X)] est indéniablement causale, compte tenu de son annotation dans la base de données IARC *TP53* et de la présentation phénotypique et familiale compatible. Pour les autres variants rares, nous constatons une très faible proportion de variants pertes de fonction (LOF) et l'absence d'arguments convergents pour établir la pathogénicité des variants faux-sens prédits délétères. A l'instar des études publiées de panels de gènes dans des populations similaires de cancers du sein [119, 165, 199, 210-216] nous retrouvons de manière très suggestive une **proportion élevée de variants affectant les voies de la réparation de l'ADN**, notamment les gènes du complexe de Fanconi et de la recombinaison homologue (*CHEK2*, *RAD51C*, *RINT1*, *WRN*). Nous constatons également un nombre de variants hétérozygotes affectant des gènes de la réparation NER ou BER, correspondant à des variants délétères en condition récessive, mais pour lesquels peut être fondée l'**hypothèse d'un allèle hypomorphe** contribuant dans un modèle oligogénique ou environnemental particulier à la survenue d'un cancer du sein précoce (*XPC*, *SLX4*, *MUTYH*, *DDB2*, *ERCC2* et *ERCC3*). La contribution des gènes du complexe de Fanconi dans la survenue de cancers héréditaires est évidente, au regard des gènes majeurs de la prédisposition héréditaire au cancer du sein et de l'ovaire que sont *FANCS/BRCA1*, *FANCD1/BRCA2*, *FANCN/PALB2*, *FANCO/RAD51C* ou encore *FANCI/BRIP1*. L'anémie de Fanconi (FA) résulte d'altérations autosomiques récessives avec altérations bi-alléliques le plus souvent hétérozygotes composites, à l'exception de FA-B lié au chromosome X et de FA-R lié à des variations dominantes négatives

de *RAD51*. La survenue de cancers précoces est intégrante de la description clinique de l'anémie de Fanconi, qui associe par ailleurs syndrome malformatif (petite taille, anomalies des membres, microcéphalie, anomalies ophtalmiques et du tractus génito-urinaire) et une insuffisance médullaire. Ces cancers sont de nature hématologique (myélodysplasie, leucémie myéloïde aigüe) mais également solide avec une prévalence importante de cancers de la sphère ORL ou génitale en relation avec l'exposition aux papillomavirus. Une étude génomique de patients de moins de 50 ans affectés de cancers de la sphère ORL révèle pour 15 individus sur 417 la présence de variants bi-alléliques sur *BRCA2*, *ERCC4*, *FANCA*, *FANCI*, *FANCM* et *SLX4*, ou mono-allélique pour *FANCB* situé sur le chromosome X [304]. Si pour certains patients, l'interprétation de ces variants bi-alléliques suggérerait une possible anémie de Fanconi, il n'est cependant pas retrouvé de corrélation avec l'âge de survenue ou la présence de signes cliniques de FA. Des formes atypiques de FA ont été décrites, notamment dans une famille remarquable par une agrégation de cancers colorectaux comportant une variation bi-allélique de *FANCD1/BRCA2* [305]. L'absence de tableau de FA-D1, pourtant de phénotype particulièrement sévère, s'explique par l'un des variants, de type faux-sens activant un site cryptique d'épissage et ne générant que partiellement un transcrit aberrant. De même est rapporté une famille affectée par une variation bi-allélique de *FANCN/PALB2* avec un lymphome de type B mais sans FA, compte tenu d'un allèle hypomorphe [306]. Ainsi, on ne peut exclure des variations bi-alléliques sur les gènes du complexe Fanconi chez des individus ne présentant pas d'anémie de Fanconi. Des enrichissements en variants appartenant aux gènes du complexe Fanconi, hétérozygotes, sont rapportés dans des populations de cancers du sein héréditaires [307, 308], avec néanmoins des niveaux de significativité discutables. Une cohorte de 83 femmes, évocatrices d'un syndrome de Li-Fraumeni en raison majoritairement d'un cancer du sein très précoce, présente pour 7 d'entre elles des variants tronquants (n=1) ou faux-sens classe 4-5 (n=6) sur des gènes appartenant au complexe de Fanconi, et pour 2 des variants faux-sens classe 4-5 sur les gènes *RECQL4* et *WRN* [309]. Des associations significatives sont rapportées entre la présence de variants hétérozygotes perte de fonction de *FANCM* et la survenue de cancers du sein précoces [303, 310, 311]. De même, un enrichissement en variants rares hétérozygotes du complexe de Fanconi est rapporté dans une cohorte de patients australiens affectés par une leucémie aigüe myéloïde [312]. Cependant dans une large étude de registre, les apparentés hétérozygotes de patients affectés de FA ne présentaient pas de risque significativement augmenté de cancers, excepté pour *FANCC* et le risque de cancer du sein (SIR=1,7) [313].

La question également critique posée par nos résultats est l'éventuelle contribution à la survenue d'un cancer, de variants rares prédits délétères, présents à l'état hétérozygote mais impliqués à l'état homozygote dans des syndromes récessifs sévères. Cette problématique d'**haplo-insuffisance** est largement illustrée dans la littérature, concernant par exemple *BLM*. Le syndrome de Bloom (OMIM: 210900) résulte de l'altération homozygote du gène *BLM*, appartenant à la famille des *RECQ-helicases*. Il associe à un syndrome malformatif une susceptibilité accrue aux cancers, notamment en raison de l'instabilité chromosomique résultant du défaut de *BLM*. Un enrichissement en variants rares hétérozygotes de *BLM* est ainsi rapporté

dans des cohortes de cancers colorectaux précoces [314], néanmoins non retrouvée dans une cohorte similaire israélienne pour une variation fondatrice tronquante [315]. De même, les porteurs hétérozygotes de variations délétères de *NBN*, gène responsable du syndrome de Nijmegen Breakage (OMIM: 251260), sont à risque accru de cancers du sein, de la prostate, du colon ou de cancers hématologiques [316]. Ces constatations sont supportées par un modèle murin *Nbn*^{+/-} avec des souris développant spontanément ces tumeurs et particulièrement susceptibles aux effets des radiations ionisantes [317]. Un enrichissement en variants rares hétérozygotes parmi *NBN*, *MRE11* et *RAD50* a également été décrit dans une cohorte de cancers du sein [140]. Enfin, les porteurs hétérozygotes de variants d'*ATM* semblent à risque accru de cancers du sein [156, 157], avec de manière identique un modèle murin supportant cette hypothèse d'haplo-insuffisance [318]. Nos observations répétées de variants rares prédits délétères, hétérozygotes, parmi les gènes *XPC*, *SLX4*, *ERCC2*, *ERCC3*, *MUTYH*, *FANCL* et *FANCM* constituent des pistes pour une possible causalité dans un modèle monogénique ou polygénique impliquant une haplo-insuffisance. Nous formulons également la perspective d'identification d'un nombre plus conséquents de variants délétères par une capture de gènes élargie à l'ensemble des composants du complexe de Fanconi. En effet, aux côtés des 21 gènes dont l'altération soit homozygote (*BRCA2*, *FANCA*, *FANCC*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *BRIP1*, *FANCL*, *FANCM*, *PALB2*, *RAD51C*, *SLX4*, *ERCC4*, *UBE2T*, *XRCC2*, *MAD2L2*, *RFWD3*), soit hémizygote lié à l'*X* pour *FANCB* soit autosomique dominant pour *FANCR/RAD51*, est associée à la survenue d'une anémie de Fanconi, d'autres gènes contribuent à cette voie complexe en tant qu'interacteurs directs ou indirects, en amont ou en aval du complexe. Ainsi, tel que revu par Ceccaldi *et al* récemment [319], on retrouve en interacteurs de *FANCM* les protéines FAAP24, FAAP16 et FAAP10. L'ubiquitylation de *FANCD2* et *FANCI*, essentielle pour l'activation du complexe et effectuée par *FANCL* et *UBE2T*, nécessite en amont *RAD18* et *PCNA*. La réversibilité de l'ubiquitylation de *FANCD2* est permise par *USP1* et *UAF1*, et requiert également *UBL5*. Des réactions de sumoylation sont également nécessaires pour la disruption du complexe *FANCD2-I*, médiées par *PIAS1* et *PIAS4*, et l'éviction du complexe des régions de l'ADN réparées requiert *RNF4* et *p97* segregase. Ainsi, sans même détailler les cascades protéiques permettant la réparation des ICL, de nombreux interacteurs sont décrits et complexifient le fonctionnement de la machinerie Fanconi. Chacun de ces acteurs constitue ainsi un potentiel candidat dans la quête de déterminants génétiques à la prédisposition aux cancers. Ainsi, dans une population de sarcomes adultes sporadiques chez des sujets de moins de 50 ans, l'essentiel des altérations identifiées (11 sur 13) concernait les gènes de la réparation de l'ADN parmi lesquels *ATM*, *BRCA2*, *ERCC4*, *FANCC*, *FANCE*, *FANCI*, *MSH6*, et *POLE* [320].

La recherche d'arguments statistiques en faveur d'un **enrichissement en variants rares** est une approche largement déployée à l'heure des séquençages massif d'exomes / génomes. La question de la population contrôle est cruciale, compte tenu des différences fondamentales d'une population à l'autre dans les fréquences alléliques. La population ExaC, même si décomposée en différentes sous-populations permettant de s'approcher de la composition ethnique des populations investiguées, reste néanmoins « polluée » puisque définie uniquement sur une origine

géographique déclarative, et par ailleurs ne comporte aucunement la garantie d'être indemne de la condition physique étudiée. La méthodologie de filtration des variants est également critique, puisque tant pour la significativité que pour la pertinence scientifique, la dépollution de tous les artéfacts, variants faux-sens de classe 1-3 et synonymes est cruciale. La filtration sur des variants rares, avec le choix d'un seuil de MAF important, permet d'augmenter la puissance mais est à définir en fonction de la prévalence de la pathologie étudiée. En effet, un seuil de MAF de 10^{-4} peut être considéré pour des pathologies génétiques rarissimes, mais est probablement trop drastique pour les cancers héréditaires, ceux-ci constituant la première maladie génétique par ordre de prévalence. Enfin, la méthodologie statistique est fondamentale, avec un niveau de maturation demeurant encore fragile s'agissant des tests d'agrégation et du choix des conditions de réalisation de ces tests (nombre de simulations, pondérations choisies...). Afin d'augmenter notre puissance, après avoir constaté qu'aucun gène ne franchissait le seuil de significativité, nous avons ainsi agrégé les variants par listes de gène. Si par cette méthode, pertinente d'un point de vue intellectuel, les *odds ratio* obtenus deviennent suggestifs d'un probable enrichissement, la **significativité n'est pas obtenue par manque probable de puissance** dès lors qu'on applique les corrections de Bonferroni. Nous avons refusé certains artifices utilisés dans la littérature par d'autres, tel que le calcul d'*odds ratios* non pas en comparant les variants LOF sur certains gènes (*DNArepairGene*) avec les gènes contrôles (*NotDNArepairGene*), mais en comparant les variants LOF sur les gènes *DNArepairGene* par rapport aux variants synonymes qu'ils comportent. Ainsi, dans l'article de Ballinger *et al* de 2016 [321], portant sur le séquençage de plus de 1000 patients adultes atteints de sarcomes, un enrichissement en variants rares classes 4/5 est rapporté significatif comparativement aux patients contrôles, en rapportant pour chaque gène impliqué le nombre de variants rares prédits pathogènes au nombre de variants synonymes portés par les mêmes gènes. Cette donnée en effet renseigne *stricto sensu* sur la mutabilité de chaque gène, mais ne constitue en aucune façon un argument en faveur d'un enrichissement chez les patients.

Partie III : recherche de variations délétères en mosaïque de TP53

Rationnel

Notre démarche nous a conduit logiquement à considérer plus avant la possible contribution de variations délétères en mosaïque pour les gènes canoniques de la prédisposition héréditaire au cancer du sein, et notamment *TP53*, considérant notre recrutement de cancers du sein avant 31 ans, de présentations sporadiques majoritairement. En effet, nous avons investigué dans la première partie la contribution de variations *de novo* dans la survenue de cancers de phénotype extrême par une approche exomique soustractive, dans une temporalité où la survenue de ces variants *de novo* a été observée à une fréquence de 1.58 par exome [232, 322]. Il convient de rappeler en effet que les variations en mosaïque correspondent à des variations *de novo*, survenant en post-zygotique dans les premières cellules embryonnaires, comme développé dans notre partie introductive. Dans notre seconde approche basée sur une approche par capture ciblée permettant une profondeur de lecture supérieure à 1000x, nous avons pu exploiter des outils tels que Varscan2 pour la détection de variants en mosaïque, et avons ainsi identifié chez une jeune femme affectée par un cancer du sein bilatéral à 27 et 34 ans sans présentation familiale particulière une variation délétère de *TP53* en mosaïque. Ces variants *de novo* en mosaïque de *TP53* ont été rapportés de manière anecdotique, tel que dans la série de Gonzalez *et al* [115] avec 7% de variations acquises détectées après analyse des parents, ou encore dans la large série de patients affectés de cancers pédiatriques (N=1120) où 50 variations délétères de *TP53* sont identifiées, comme attendu, dans une cohorte de tumeurs pédiatriques [323]. Trois de ces variations de *TP53* sont identifiées à l'état de mosaïque, dans un contexte de leucémie aigüe lymphoblastique. Auparavant, Behjati *et al* [274] avait déjà rapporté chez un enfant atteint successivement de sarcome puis de neuroblastome une variation en mosaïque de *TP53*.

Nous avons de ce fait considéré l'opportunité d'une recherche extensive de ces événements en mosaïque, en ciblant particulièrement *TP53* compte tenu de notre recrutement de tumeurs mammaires précoces et sporadiques avant 31 ans et de cancers pédiatriques, lié au positionnement national du Laboratoire de Génétique du CHU de Rouen. Au-delà de l'identification de l'altération causale, la nature en mosaïque d'une telle altération est particulièrement informative concernant le conseil génétique. En effet, s'agissant d'une variation apparue en post-zygotique, il peut dès lors être communiqué aux parents et aux fratries l'absence de risque génétique et ainsi apporter la réassurance et limiter les lourdes surveillances proposées aux apparentés dans ces contextes sévères. En revanche, dans la limite actuelle des capacités d'identification des tissus concernés par ces variations en mosaïque et notamment de l'atteinte des lignées gonadiques, le risque de transmission à la descendance des sujets atteints ne peut être correctement apprécié.

Matériel et Méthode

Patients

Dans cette approche, la recherche de variations de *TP53* en mosaïque a été conduite dans 3 cohortes indépendantes :

- **Cohorte A** (N=308) : patients séquencés par la capture 201 gènes, comprenant l'ensemble des sous-populations de cette étude à savoir cancers du sein <31 ans (N=105), cancers de l'ovaire <41 ans (N=13), cancers digestifs <31 ans (N=84), cancers pédiatriques (N=92 avec rhabdomyosarcome (44%), corticosurrénalome (34%) et tumeurs des plexus choroïdiens (23%)) ou cancers multiples de phénotype extrême (N=14). Parmi ces différentes sous-population, j'ai personnellement contribué à la cohorte des cancers du sein.
- **Cohorte B** : patients adultes ou enfants non apparentés, répondant aux critères de Chompret, adressés au Laboratoire de Génétique du CHU de Rouen pour analyse *TP53* et porteurs d'une variation délétère identifiée (N=328). Cette cohorte a été enrichie par 108 patients fortement évocateurs d'un syndrome de Li-Fraumeni en raison d'un corticosurrénalome (N=55), d'une tumeur des plexus choroïdes (N=21), d'un cancer du sein précoce avant 31 ans (N=31) ou de tumeurs primitives multiples (N=1), mais sans altération identifiée par analyse Sanger et QMPSF. Cette cohorte a été prise en charge par ma collègue Mme RENAUX-PETEL.
- **Cohorte C** : patients présentant un cancer du sein précoce avant 31 ans, quel que soit l'histoire familiale et le phénotype tumoral, sans altération identifiée par les techniques diagnostiques conventionnelles sur les gènes canoniques de la prédisposition héréditaire au cancer du sein comprenant *BRCA1*, *BRCA2* et *TP53* (N=63). J'ai personnellement constitué et analysé cette cohorte.

Pour l'ensemble des patients, nous avons pu disposer du consentement originel autorisant les explorations moléculaires de *TP53* dans le contexte de la pathologie explorée, à la fois à finalité diagnostique mais également à finalité de recherche.

Analyse de *TP53* par séquençage Sanger et Quantitative Multiplex PCR of Short fluorescent Fragments

Les patients de la cohorte B ont bénéficié dans un premier temps d'une analyse moléculaire conventionnelle de *TP53*. L'ADN leucocytaire extrait d'un prélèvement sanguin a été amplifié par réaction PCR pour les 11 régions exoniques de *TP53*. L'analyse par séquençage a

été effectuée par méthode de Sanger, sur séquenceur *3130xl Genetic Analyser* (Applied Biosystems) avec recherche de variations ponctuelles par le logiciel *Variant Reporter software V1.1* (Applied Biosystems). La recherche de réarrangement de grande taille a été menée par la méthode *Quantitative Multiplex PCR of Short fluorescent Fragments* (QMPSF) selon la méthodologie développée par l'unité Inserm 1245 [324, 325]. L'analyse a été conduite par le logiciel *Genescan* (PE Applied Biosystems) après électrophorèse sur séquenceur *3130xl Genetic Analyser* (Applied Biosystems).

Confirmation de la filiation des individus

L'analyse portant sur la recherche de variants *de novo* en mosaïque de *TP53*, l'approche idéalement doit s'appuyer sur une certitude concernant l'identité des parents du cas index. Ainsi, pour les patients dont nous disposons des ADN parentaux (cohorte B uniquement), une analyse des profils microsatellitaires a été conduite en s'appuyant sur les loci D1S439 (AFM225xe11), D9S1784 (AFMa136xa5), D14S986 (AFMa184xa5), D19S913 (AFMb301xc9), D17S960, D17S1353 et D17S1844.

Préparation des librairies d'ADN, séquençage de nouvelle génération et analyses bio informatiques

Nous avons procédé à des méthodologies différentes pour les 3 cohortes, correspondantes aux évolutions technologiques et aux optimisations internes conduites sur la période de l'étude.

Ainsi, pour la cohorte A correspondant aux patients analysés par la capture ciblée 201 gènes, la préparation des librairies a été effectuées par la technologie QXT (*SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing*) tel que décrite dans notre deuxième partie. Le séquençage a été réalisé sur le séquenceur NextSeq en *paired-end* 2x150pb.

La cohorte B correspondant aux patients référés pour analyse moléculaire de *TP53* ont bénéficié également d'une préparation des librairies selon la technologie QXT. La capture utilisée ciblait bien entendu *TP53*, mais également d'autres gènes impliqués dans le panel oncologique orienté cancers colorectaux. Cette capture correspond en effet à celle validée par le Laboratoire de Génétique du CHU de Rouen pour l'exploration des suspicions de prédisposition héréditaires aux cancers digestifs. Le séquençage a été réalisé sur les plateformes MiSeq ou NextSeq500 en *paired-end* 2x150pb.

La cohorte C composée de patients atteints exclusivement de cancers du sein précoce (<31 ans) a bénéficié d'une préparation des librairies d'ADN assistée par la station de travail Sciclone© G3 NGS de Perkin-Elmer, selon la méthodologie QXT (*SureSelect^{QXT} Target Enrichment*

for *Illumina Multiplexed Sequencing*). Le séquençage a été réalisé sur le séquenceur NextSeq en *paired-end* 2x150pb.

Les analyses bio informatiques ont suivi le pipeline tel que décrit dans la partie 2, composée du logiciel CASAVA version 1.8 d'Illumina pour le démultiplexage, la suite BWA - Picard - GATK pour l'alignement et le *base calling*, GATK HaplotypeCaller et VarScan2 pour le *variant calling*. L'annotation a été réalisée par la suite Alamut Batch (*Interactive Biosoftware*).

Confirmation des variants en mosaïque par SNaPshot™

Afin de limiter le risque de faux-positif du NGS (artéfact de PCR), une confirmation par une technique orthogonale est nécessaire et recommandée. S'agissant de variants en mosaïque, leur détectabilité par méthode de Sanger est compromise dès lors que la balance allélique est faible (<20%). Nous avons dès lors utilisé la méthodologie d'extension de *primers* (*ABI PRISM SNaPshot™ Multiplex kit, Applied Biosystems*) selon les recommandations du fabricant : après une PCR classique des séquences comportant la variation d'intérêt, les *primers* et dNTP en excès sont éliminés par un traitement à la *shrimp alkaline phosphatase* (SAP) et *Exonuclease I* (Exo I). La réaction d'extension d'amorce est alors réalisée selon le principe de ddNTP marqués par un fluorochrome se polymérisant à partir de l'extrémité 3'OH de l'amorce, de manière complémentaire à la matrice ADN amplifiée. Cette liaison interrompt l'extension, permettant ainsi de révéler une altération ponctuelle de manière ultra-sensible par le choix d'une amorce précédant immédiatement la position mutée. Les produits de réaction sont analysés après électrophorèse sur séquenceur 3130xl *Genetic Analyser* (*Applied Biosystems*) par le logiciel GeneMapper V.4.1.

Résultats

Cohorte A

Parmi la cohorte A correspondant aux 308 patients inclus dans l'étude par capture des 201 gènes, se distinguent :

- 105 patients atteints de cancer du sein précoce avant 31 ans, incluant les 30 patients précédemment décrits dans la Partie II, caractérisés par une présentation sporadique dans 72% des cas. Les autres patients inclus par le Laboratoire de Caen (Inserm U1245, D. Vaur) sont issus du recrutement des patients adressés au Laboratoire pour une analyse HBOC dans un contexte de suspicion de prédisposition héréditaire au cancer du sein et de l'ovaire, expliquant une plus grande proportion de patients avec histoire familiale.
- 13 patients atteints de cancer de l'ovaire précoce avant 41 ans, également avec présentation familiale.

- 84 patients présentant une tumeur colorectale avant 31 ans, issus principalement du recrutement de l'étude précédemment publiée par l'équipe Inserm U1245 [326] s'intéressant au génotype de SNPs associés à un risque accru de cancer colorectal. Ces patients devaient avoir bénéficié d'un diagnostic d'exclusion pour les formes connues de prédisposition héréditaire au cancer colorectal (HNPCC, PAF et Hamartomatoses) et pouvaient être inclus sur la base d'une histoire familiale (cancer colorectal chez un apparenté au 1^{er} ou 2^{ème} degré dont l'un avant 61 ans) ou de la précocité de survenue (avant 41 ans) du cas index ou de la multiplicité des cancers colorectaux chez un même individu. Nous avons restreint cette série aux cancers colorectaux survenus avant 31 ans chez le cas index.
- 92 patients atteints de cancers pédiatriques, et majoritairement par les tumeurs emblématiques du syndrome de Li-Fraumeni : rhabdomyosarcome (44%), corticosurrénalome (34%) et tumeurs des plexus choroïdiens (23%). Ces patients avaient bénéficié antérieurement d'une analyse diagnostique conventionnelle du gène *TP53* sans altération identifiée.
- 14 patients adultes atteints de cancers primitifs multiples, remarquables par cette agrégation individuelle.

L'ensemble des patients a été capturé par le panel de 201 gènes. Une visualisation directe des BAM focalisée sur *TP53* a été réalisée pour l'ensemble des patients. Une analyse par VarScan 2 a été conduite, globale sur les 201 gènes mais avec également un focus particulier sur *TP53*. Seule la patiente #24 s'est avérée porteuse d'un variant *de novo* en mosaïque telle que décrit dans la partie II, soit la détection d'une variation en mosaïque sur 308 patients (0,3%).

Cohorte B

Nous avons repris l'ensemble des 328 cas index de la collection Li-Fraumeni du Laboratoire de Génétique de Rouen présentant une altération délétère de *TP53* détectée par méthode de Sanger, pour lesquels nous disposons des ADN des parents. Le séquençage également selon la méthode Sanger des parents a permis ainsi de constater que pour 40/328 individus (12%), la variation de *TP53* identifiée chez les cas index était apparue *de novo*. Cette conclusion a bien sûr été étayée par l'analyse microsatellitaire permettant de s'assurer de la filiation de chacun de ces cas index. Pour 28 trios parents - enfant, la qualité de l'ADN a permis d'envisager une analyse NGS à forte profondeur (400 - 1000x) à même de valider l'absence de la variation *TP53* chez les deux parents. Les caractéristiques de ces 40 patients, et des variations délétères de *TP53* identifiées *de novo* sont résumées dans le tableau 12. On note bien entendu une proportion majoritaire de cas pédiatriques comme attendu, mais néanmoins 17 patientes présentait un cancer du sein précoce avant 31 ans, confirmant la prévalence élevée des cancers du sein dans le syndrome de Li-Fraumeni. On note également la multiplicité des tumeurs primitives chez les cas index (28/40), marquant ainsi la sévérité de ce syndrome. Il n'a pas été retrouvé dans notre cohorte d'effet de l'âge paternel pouvant être à l'origine de ces variations *de*

novo, avec un âge moyen lors de la conception de 31 ans pour les pères, et de 28 ans pour les mères, correspondant aux âges moyens observés dans la population française.

Famille	Genre	Phénotype (âge en années)	Mutation	% de reads mutants
F192	M	RMS (1) ; OS (11) ; OS (19)	c.844C>T p.(Arg282Trp)	55% (424/767)
F70	F	LMS (39) ; BC (42)	c.476C>A. p.(Ala159Asp)	52% (489/932)
F194	M	BL (2) ; OS (16) ; GA (17)	c.761T>A p.(Ile254Asn)	52% (552/1059)
F304	F	ACC (21)	c.451C>G p.(Pro151Ala)	51% (411/799)
F256	F	ALL (16)	c.743G>A p.(Arg248Gln)	49% (389/788)
F267	F	OS (16) ; BC (28)	c.743G>A p.(Arg248Gln)	49% (517/1048)
F234	M	RMS (2) ; OS (12)	c.518T>C p.(Val173Ala)	49% (524/1067)
F216	M	WT (2) ; MB (7)	c.378C>G p.(Tyr126*)	49% (354/724)
F279	F	ACC (5)	c.527G>A p.(Cys176Tyr)	49% (475/974)
F215	F	ACC (25) ; BC (28)	c.844C>T p.(Arg282Trp)	48% (327/678)
F66	F	OS (18) ; BBC (34/35)	c.310C>T p.(Gln104*)	48% (360/750)
F270	M	CBC (33) ; STM (40)	c.613T>G. p.(Tyr205Asp)	48% (307/643)
F60	F	BBC (25/35)	c.1009C>T. p.(Arg337Cys)	47% (482/1015)
F50	F	RMS (1) ; OS (2)	c.814G>A p.(Val272Met)	47% (243/517)
F143	F	RMS (1) ; RMS (2)	c.672+1G>T p.?	47% (337/717)
F235	M	CPC (0,5)	c.365_366del p.(Val122Aspfs*26)	46% (311/672)
F174	F	OS (11)	c.919+1G>T p.?	43% (284/667)
F47	F	BC (26) ; GB (27)	c.393_395del p.(Asn131del)	42% (321/761)
F83	M	STM (38) ; STM (44)	c.329G>C p.(Arg110Pro)	42% (352/839)
F171	F	CPC (1)	c.524G>A p.(Arg175His)	41% (354/865)
F38	M	OS (19) ; CS (31)	c.842A>T p.(Asp281Val)	41% (266/655)

Famille	Genre	Phénotype (âge en années)	Mutation	% de reads mutants
F148	F	CPC (17) ; BS (24)	c.845G>C p.(Arg282Pro)	38% (225/585)
F199	F	BBC (30)	c.782+1G>A p.?	38% (452/1200)
F287	F	HD (16) ; AML (24) ; BC (25)	c.428T>C p.(Val143Ala)	34% (302/877)
F266	M	GB (3)	c.731G>A p.(Gly244Asp)	34% (363/1077)
F127	F	BBC (29)	c.913_916del p.(Lys305Glufs*39)	30% (175/575)
F186	F	BBC (25)	c.632_641del p.(Thr211Ilefs*33)	28% (202/733)
F277	M	MB (9) ; OS (13)	c.96+31_97-32del p.?	14% (76/550)
F7	F	STM (2) ; ACC (6) ; OS (14) ; BBC (24/25)	c.535C>T p.(His179Tyr)	ND
F62	F	ACC (1) ; RMS (1) ; OS (17)	c.818G>T p.(Arg273Leu)	ND
F96	M	CPC (4)	c.733G>A p.(Gly245Ser)	ND
F115	F	BBC (27)	c.390_426del p.(Asn131Cysfs*27)	ND
F118	M	MB (10)	c.376-2A>G p.?	ND
F155	F	BC (52)	c.524G>A p.(Arg175His)	ND
F180	M	STM (46) ; AML (48)	c.743G>A p.(Arg248Gln)	ND
F245	F	CRC (24) ; BBC (31/34)	c.374C>A p.(Thr125Lys)	ND
F321	F	BC (29) ; OS (35)	c.681dup p.(Asp228*)	ND
F327	M	Indemne (33)	c.535C>G p.(His179Asp)	ND
F342	F	BBC (26)	c.817C>G p.(Arg273Gly)	ND
F343	M	ACC (1)	c.844C>T p.(Arg282Trp)	ND

Tableau 12 : Génotype et phénotypes associés des variations *de novo* de *TP53* identifiées

Abbreviations :

RMS : Rhabdomyosarcome ; OS : Ostéosarcome ; LMS : Léiomyosarcome ; BC : Cancer du sein ; BBC : Cancer du sein bilatéral ; BL : Lymphome de Burkitt ; GA : Adénocarcinome Gastrique ; WT : Tumeur de Wilms ; MB : Médulloblastome ; ACC : Corticosurrénalome ; CBC : Carcinome Basocellulaire ; CPC : Carcinome des Plexus Choroïdes ; GB : Glioblastome ; CS : Chondrosarcome ; BS : Sarcome mammaire ; HD : Maladie de Hodgkins ; AML : Leucémie Aigüe Myéloïde ; CRC : Cancer Colorectal

Par une analyse attentive des électrophorégrammes des analyses Sanger, nous avons identifié en outre deux variations additionnelles de *TP53* présentes à l'état de mosaïque. Dans la famille F281, dont le cas index présentait la variation hétérozygote c.814G>A, p.(Val272Met) dans un contexte de médulloblastome sporadique à 3 ans, l'analyse du père indemne a révélé la présence discrète de la variation sur l'analyse Sanger, avec un pic d'amplitude comparable au bruit de fond. Afin de confirmer cette hypothèse d'un variant délétère de *TP53* en mosaïque impliquant la lignée gonadique chez le père, nous avons ré-analysé l'ADN de ce dernier par SNaPshot et NGS, retrouvant la variation à hauteur de 5% (43 reads mutants sur 866) (Figure 19).

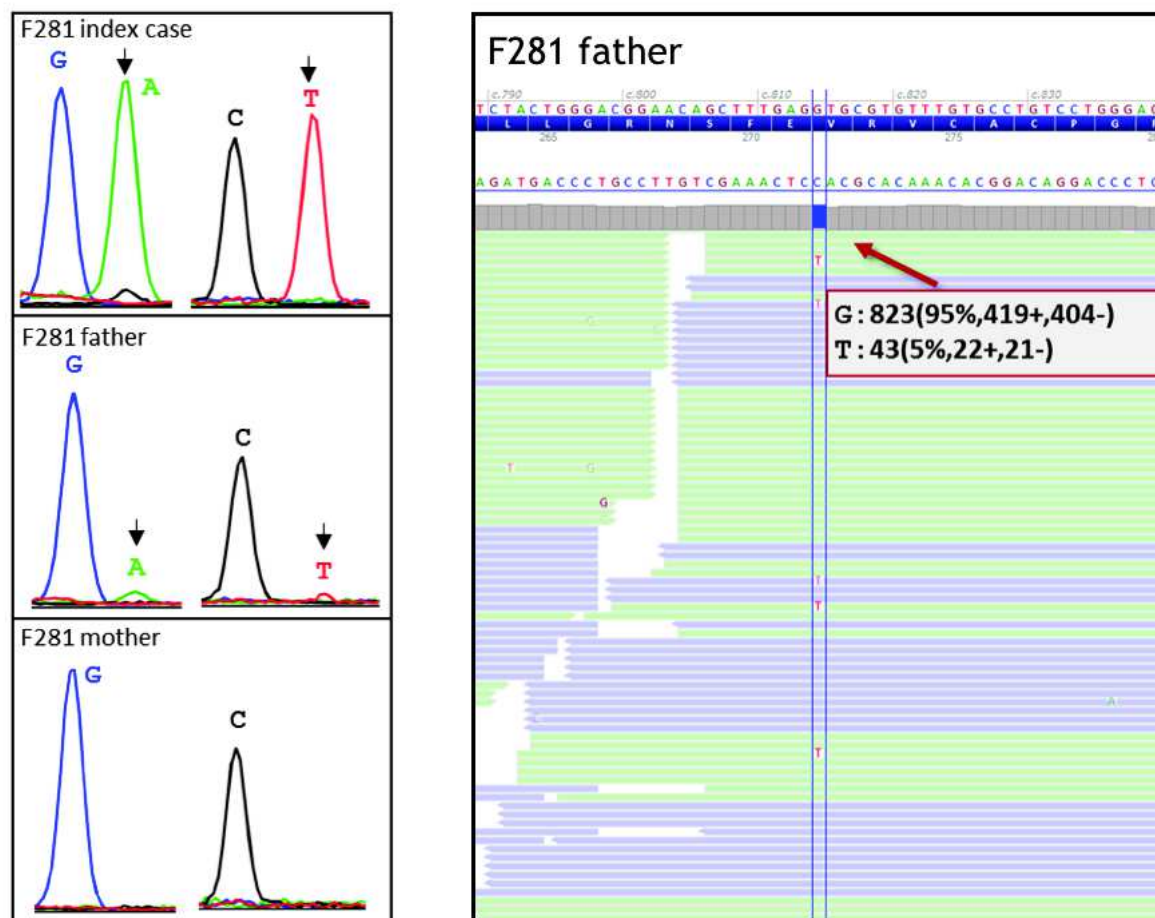


Figure 19 : Analyse par SNaPshot (séquences Forward et Reverse) et analyse NGS du trio F281. La flèche indique la variation *TP53* c.814G>A.

De même, une variation en mosaïque de *TP53* dans l'exon 7 (c.[722C=>T], p.[Ser241=>(Ser241Phe)]) est retrouvée à état de trace sur le séquençage Sanger d'un enfant de 4 mois (F315) atteint de corticosurrénalome, estimée à 17% par SNaPshot et NGS (Figure 20).

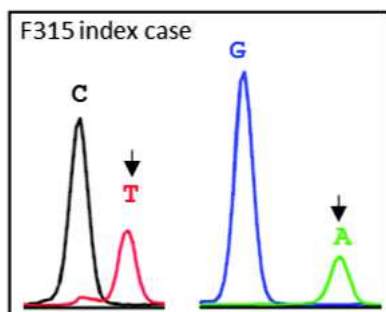


Figure 20 : Analyse par SNaPshot (Forward et Reverse) et NGS du cas index F315. La flèche indique la variation *TP53* c.[722C=>T]

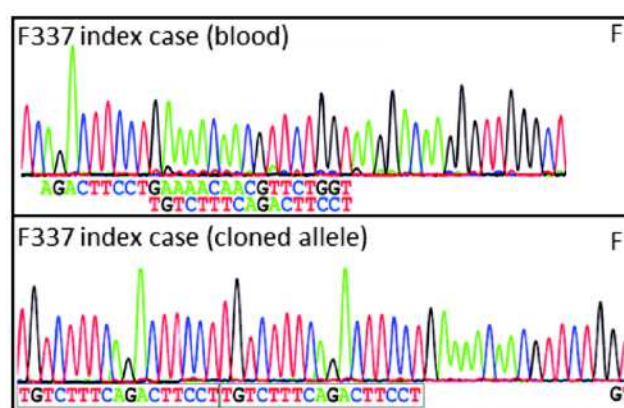
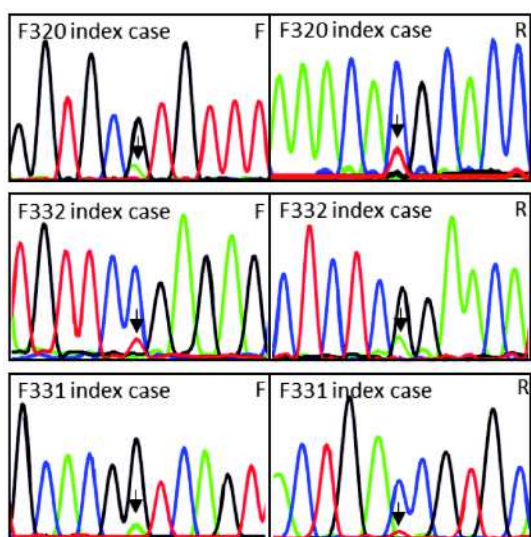
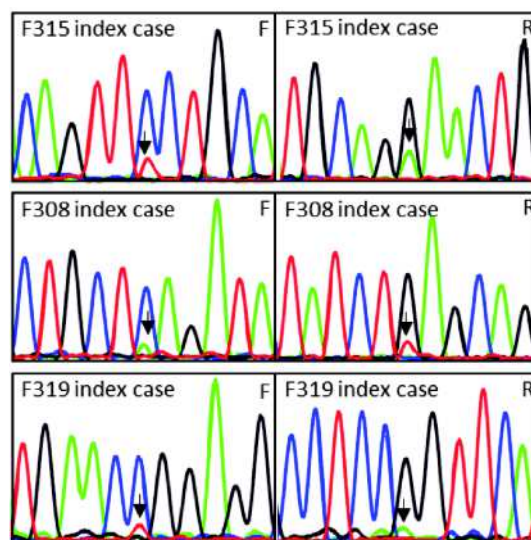
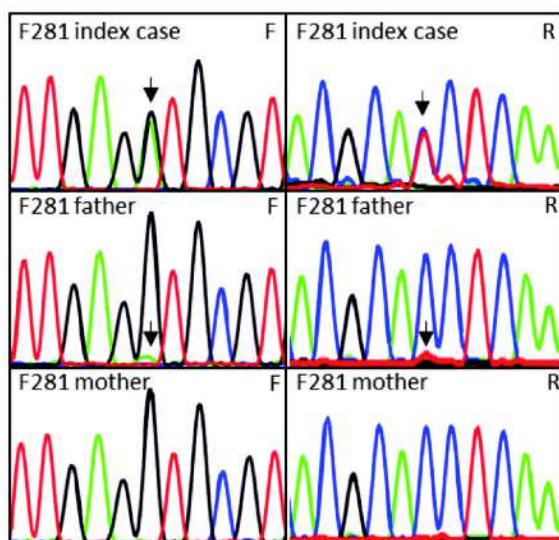


Figure 21 : Variations *TP53* en mosaïque sur électrophorégramme Sanger (Séquences Forward et Reverse). Les flèches indiquent les variations *TP53*.

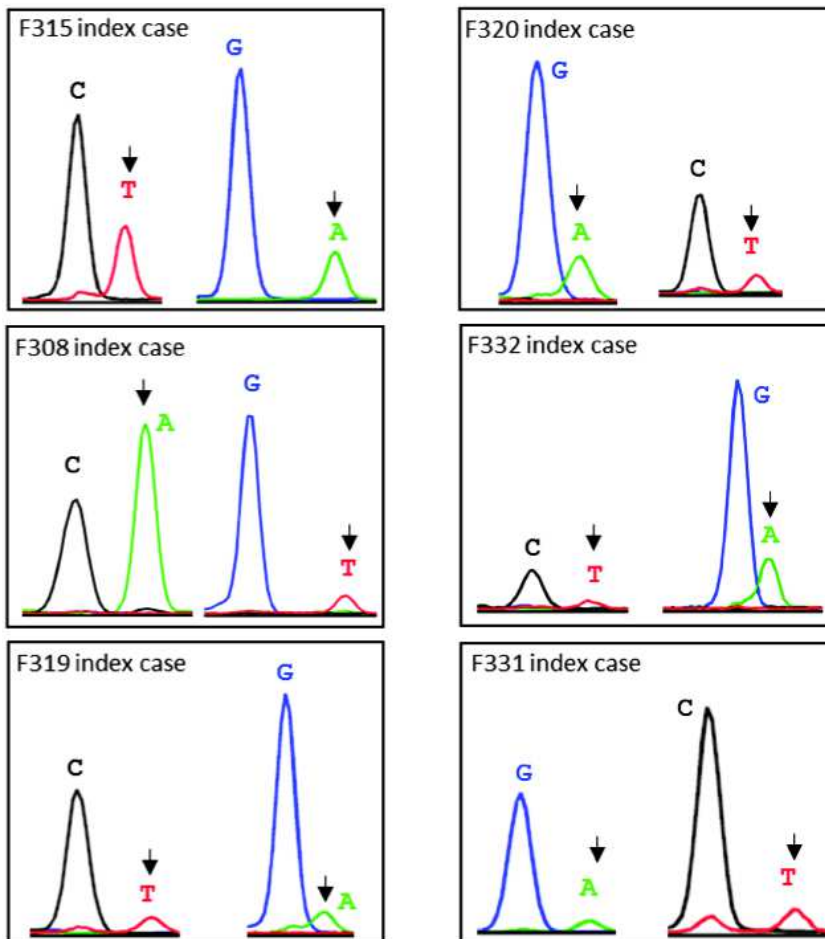


Figure 22 : Confirmation des mosaïques par SNaPshot (Séquences Forward et Reverse). Les flèches indiquent les variations *TP53*.

Compte tenu de ces deux observations soulignant la proportion importante de variations *de novo* de *TP53* (12%) dans une cohorte de patients répondant aux critères de Chompret, associées à deux variants en mosaïque, nous avons étendu cette cohorte par l'analyse de 108 patients supplémentaires atteints des pathologies emblématiques du syndrome de Li-Fraumeni que sont les corticosurrénalomes, les tumeurs des plexus choroïdes et les cancers du sein précoces. L'analyse du gène *TP53* par NGS de 49 enfants et 6 adultes atteints de corticosurrénalomes permet ainsi l'identification de deux variations en mosaïques supplémentaires (F308 et F337, tableau 13). De même, deux variants en mosaïques additionnelles de *TP53* sont identifiés parmi une cohorte de 21 enfants atteints de tumeurs des plexus choroïdes (F319 et F320, tableau 13). Parmi 31 patientes atteintes de cancers du sein précoce, incluant la patiente F332 analysée par la capture ciblée 201 gènes, nous n'avons pas identifié de variation en mosaïque supplémentaire. Enfin, une patiente (F331), atteinte de tumeurs primitives multiples très suggestives d'altérations de *TP53* (ostéosarcome de la base du crâne, cancer du sein *in situ* et sarcome mammaire), présente par analyse NGS une variation en mosaïque sur 7% des *reads* (tableau 13). L'ensemble de ces variants a été confirmé par NGS, SNaPshot et relecture *a posteriori* des électrophorégrammes des analyses Sanger (Figures 21 et 22). Lorsque l'analyse de l'ADN tumoral a été possible (F315, F308 et F331), nous avons pu montrer une inversion des fréquences alléliques traduisant la perte de l'allèle sauvage au niveau tumoral (perte

d'hétérozygotie) témoignant de la pathogénicité des variations identifiées (Figure 23).

Cohorte C

Afin d'explorer davantage la possible contribution de variants en mosaïque de *TP53* dans la survenue de cancers du sein précoces, nous avons étendu les séries précédentes par le recrutement et l'analyse de 63 patientes supplémentaires.

Ces patientes avaient fait l'objet antérieurement d'une analyse moléculaire diagnostique conventionnelle dans le contexte de la survenue d'un cancer du sein avant 36 ans, soit la recherche d'altérations de *BRCA1*, *BRCA2*, *PALB2* et *TP53*. Ces analyses étaient négatives. Ces patientes ne présentaient pas de contexte familial particulier, les critères de Chompret se résument exclusivement à la survenue d'un cancer du sein précoce. Sur les 48 patientes dont le phénotype tumoral était détaillé, 19 présentaient une sur-expression de HER2.

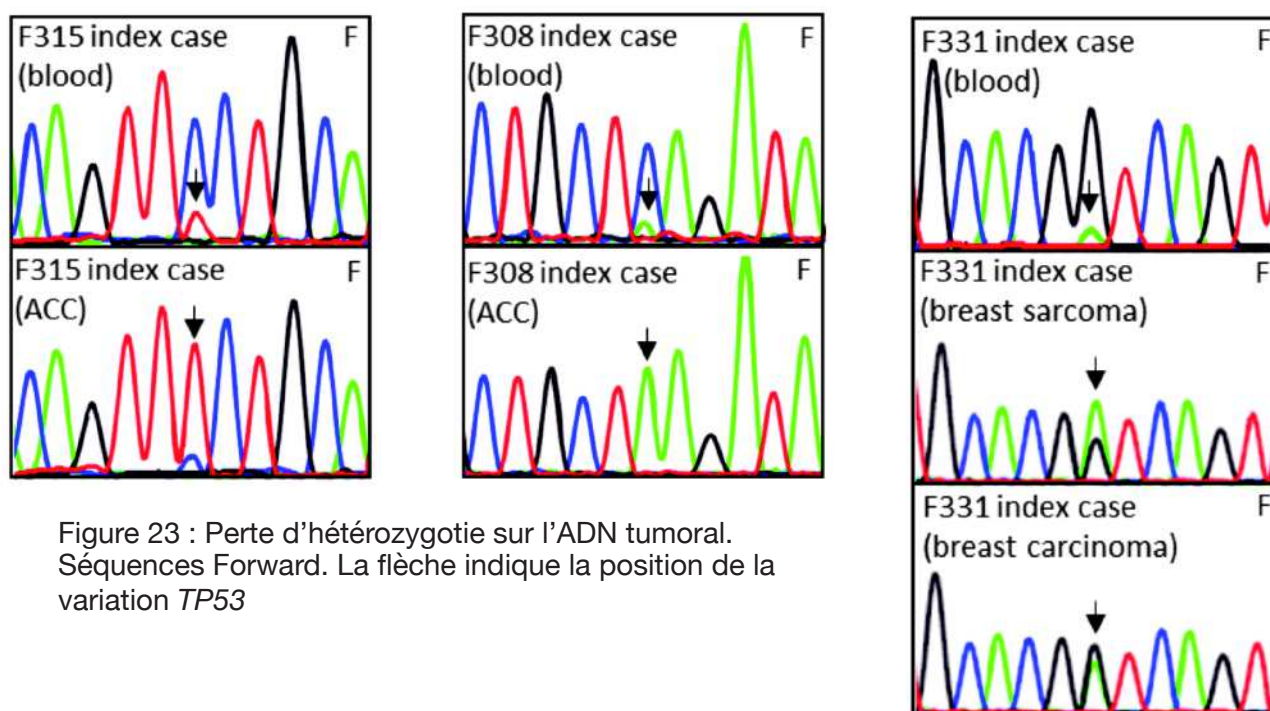


Figure 23 : Perte d'hétérozygotie sur l'ADN tumoral. Séquences Forward. La flèche indique la position de la variation *TP53*

Afin d'augmenter la capacité de détection d'évènements en mosaïque, nous avons utilisé pour cette analyse une capture conçue exclusivement pour *TP53*, incluant l'ensemble des séquences exoniques et introniques de *TP53*, les séquences 5' et 3'UTR et promotrices, les gènes situés en amont et aval ainsi que des polymorphismes impliqués en tant que facteurs modificateurs.

Deux *runs* de séquençage ont été effectués, avec des données de qualité robustes : On *Target* de 25% comme attendu dans une approche de capture très ciblée à grande profondeur, profondeur moyenne de 1500x et >90% des positions couvertes à plus de 1000x, Q-score supérieur à 30 pour 79% des données.

L'analyse bio informatique, conduite par le double pipeline HaploTypeCaller et VarScan 2 pour le *variant calling*, n'a pas permis d'identifier de variations de *TP53* en particulier en mosaïque, après élimination des artéfacts par analyse des BAM.

Discussion

Nous avons étayé, par cette analyse extensive de cohortes de patients suggestifs de syndrome de Li-Fraumeni, l'existence et la relative fréquence de variations délétères *de novo* de *TP53*, avec 48 variants *de novo* identifiés sur une série totale de 807 patients. Même si les cancers du sein précoces et les cancers colorectaux précoces peuvent s'intégrer dans le spectre tumoral du syndrome de Li-Fraumeni, nous avons constaté l'essentiel de ces variations *de novo* dans des situations de cancers pédiatriques, ou de tumeurs primitives multiples très évocatrices d'une altération de *TP53*. Dès lors, nous estimons plus probable une fréquence de ces variations *de novo* de *TP53* de l'ordre de 14%, si l'on restreint le dénominateur à ces situations plus évocatrices (Cohorte B). Cette fréquence est en cohérence avec la littérature, puisque Gonzalez *et al* [115] ont rapporté un taux de variation *de novo* de *TP53* dans 7%. Cette fréquence est néanmoins possiblement sous-estimée dans notre étude, puisque nous ne disposons pas systématiquement de l'ADN parental. En se fondant sur l'anamnèse familiale, nous pouvons ainsi suspecter une origine *de novo* pour 80 cas index supplémentaires, sans toutefois exclure que l'absence d'antécédent familiaux ne traduise une pénétrante incomplète et/ou des facteurs modificateurs.

Parmi ces variants *de novo*, nous avons identifié 8 altérations en mosaïques, apportant ainsi la contribution à ce jour la plus importante grâce à une collection importante de trios ADN cas index/ ADN parentaux et l'utilisation du NGS dans une cohorte de patients affectés de tumeurs emblématiques du syndrome de Li-Fraumeni. Ces variations sont pour la plupart indétectables par séquençage Sanger, ou ne passeraient pas les filtres bio informatiques si les profondeurs de lecture / les pipelines d'analyse bio informatiques ne prenaient pas en compte cette hypothèse. A l'inverse, dans cette quête de variations en mosaïque de *TP53*, différents travaux ont souligné le risque d'artéfacts liés à la présence d'ADN tumoral circulant (variants en mosaïque d'origine somatique), liés à une chimiothérapie en cours à l'origine d'altérations en mosaïque induite par les cytotoxiques dans les lignées hématopoïétiques [327] ou encore l'apparition de sous-clones hématopoïétiques porteurs de variations de *TP53* au cours du vieillissement des individus [328]. Nous avons pu, pour une fraction de nos individus porteurs de variants en mosaïque de *TP53*, analyser la tumeur et constater une perte d'hétérozygotie impliquant l'allèle sauvage. Nous pouvons ainsi, pour 3 des 8 individus, affirmer la causalité de cette variation en mosaïque dans la survenue des tumeurs les affectant. Nous avons pu exclure également chez la patiente F337 le caractère secondaire à la chimiothérapie de la variation identifiée, cette dernière n'ayant pas reçu de chimiothérapie. Idéalement néanmoins, ces variations en mosaïques nécessiteraient une confirmation par analyse ciblée du tissu non tumoral correspondant au tissu impliqué dans la cancérogenèse, ou à défaut, l'analyse de tissus autres

qu'hématopoïétiques mais facilement prélevables tels que fibroblastes d'origine salivaire ou cutanée, cellules urinaires... Dans le cadre du diagnostic de la polypose adénomateuse familiale, Spier *et al* [329] ont ainsi montré chez des sujets fortement évocateurs la présence sous forme de mosaïque de variation délétère d'APC, détectée uniquement par séquençage du tissu tumoral et péritumoral. Cette approche est envisagée pour la cohorte C, où nous collectons actuellement les échantillons tumoraux afin d'analyser l'ADN du tissu mammaire sain et rechercher d'éventuelles variations en mosaïque non détectées sur l'ADN leucocytaire.

L'identification de variants *de novo* de *TP53* est cruciale pour la prise en charge individuelle et familiale. En effet, comme indiqué précédemment, les sujets affectés par une variation délétère de *TP53* présentent un risque théorique de cancers de spectre large impliquant des programmes de surveillance lourds mais bénéfiques basés notamment sur des examens non irradiant tels que l'IRM corps entier ou l'échographie, répétés tout au long de la vie [125-127]. Ces programmes impliquent également des chirurgies prophylactiques mammaires chez les jeunes femmes. Enfin, comme évoqué en introduction, la génotoxicité particulière des traitements de radiothérapie ou de chimiothérapie chez les individus porteurs de ces altérations *TP53* conduisent à intégrer cette donnée dans les programmes personnalisés de soins, notamment en préférant la réalisation de mastectomie à un traitement conservateur suivi de radiothérapie, ou à favoriser des chimiothérapies à base de taxanes moins génotoxiques que les anthracyclines dans le domaine de la cancérologie mammaire. Le caractère *de novo* de ces altérations permet ainsi de restreindre ces mesures lourdes au seul individu atteint et éventuellement à sa descendance, tout en permettant de rassurer et d'exclure de ces programmes de surveillance les parents / frères et soeurs et autres apparentés. Dans le cas de variations *de novo* en mosaïque, l'identification des tissus concernés chez le sujet ouvrirait des perspectives de plus grande personnalisation des soins, en ne retenant au sein du protocole de surveillance dit de Toronto que les examens pertinents au regard des tissus impliqués par la mosaïque. L'identification d'une implication des tissus gonadiques légitimerait également ou non la prise en charge de la descendance.

Ainsi, nous démontrons que les altérations du gène *TP53* demeurent un déterminant essentiel dans les présentations tumorales précoces (cancers pédiatriques, cancers du sein de la femme très jeune) et dans les présentations tumorales multiples, et ce indépendamment de l'histoire familiale puisqu'un certain nombre de ces altérations surviennent *de novo*. Ce dernier aspect légitime particulièrement l'analyse de *TP53* sur le seul critère de Chompret d'un cancer du sein avant 36 ans. En outre, les techniques d'analyse doivent impérativement migrer vers des solutions d'analyse à grande profondeur type NGS avec des pipelines bio informatiques capables de détecter ces altérations en déséquilibre allélique, tout en se prémunissant contre les situations artéfactuelles d'ordre technologiques ou cliniques (effets de la chimiothérapie, de l'âge, ADN tumoral circulant).

Famille	Genre	Phénotype (âge en années)	Mutation	% de reads mutants	Analyse tumorale
F281	M	Indemne, 42 ans	c.[814G=>A] p.[Val272=/(Val272Met)]	5% (43/866)	-
F315	M	ACC (0,3)	c.[722C=>T] p.[Ser241=/(Ser241Phe)]	17% (181/1060)	LOH
F308	F	ACC (0,7)	c.[548C=>A] p.[Ser183=/(Ser183*)]	17% (145/832)	LOH
F337	F	ACC (1)	c.[75-10_81=dup] p.[Glu28=/(Glu28Cysfs*22)]	4% (25/571)	ND
F319	M	CPC (2)	c.[742C=>T] p.[Arg248=/(Arg248Trp)]	14% (263/1928)	ND
F320	M	CPP atypique (0,5)	c.[818G=>A] p.[Arg273=/(Arg273His)]	19% (359/1902)	ND
F331	F	OS (12) ; BC (35) ; BS (35)	c.[375+1G=>A] p.[=?]	7% (40/551)	LOH

Tableau 13 : Variations en mosaïque de *TP53* identifiées

Abbreviations :

OS : Ostéosarcome ; BC : Cancer du sein ACC : Corticosurrénalome ; CPC : Carcinome des Plexus Choroïdes ;
CPP : Papillome des plexus choroïdes ; BS (Sarcome mammaire) ; LOH : *loss of heterozygoty* ; ND : non déterminé

Article : Contribution of de novo and mosaic TP53 mutations to Li-Fraumeni syndrome.

Renaux-Petel M., Charbonnier F., **Théry JC**, Fermey P., Lienard G., Bou J., Coutant S., Vezain M., Kasper E., Fourneaux S., Manase S., Blanluet M., Leheup B., Mansuy L., Champigneulle J., Chappé C., Longy M., Sévenet N., Bressac-de Paillerets b., Guerrini-Rousseau L., Brugières L., Caron O., Sabourin JC, Tournier I., Baert-Desurmont S., Frebourg T., Bougeard G.

J Med Genet 2018;55:173-180.

Discussion Générale

Nos approches ont tenté d'implémenter les aspects technologiques et conceptuels émergents dans le diagnostic moléculaire des formes suspectées héréditaires de cancers du sein et de l'ovaire, en utilisant l'extrême précocité comme marqueur du caractère héréditaire. L'approche par exomes soustractifs dans des trios cas index / parents nous a permis de démontrer la pathogénicité d'un **variant de novo affectant le gène *INHBA*, et la contribution de cette voie biologique dans le cancer de l'ovaire**. Nous avons cependant échoué à renouveler cette démonstration dans des trios affectés par des cancers du sein précoces. Bien au-delà d'un phénotype différent, nous nous sommes heurtés à l'extrême difficulté de démontrer une pathogénicité lorsque les variations *de novo* impliquent des gènes peu connus, ou dont le lien avec la carcinogenèse est ténu. Même si le caractère *de novo* d'une variation identifiée chez un sujet atteint, issu de parents indemnes, conduit à suspecter un caractère causal, la démonstration de cette causalité implique notamment des modèles expérimentaux souvent complexes pour des gènes peu connus ou des stratégies de récurrence se heurtant au concept de variations privées.

L'approche exomique comparative dans cette famille remarquable par la survenue précoce de cancers du sein et du pancréas n'a pas permis non plus l'identification d'une variation délétère, renforçant la notion d'**altérations cryptiques** que les avancées technologiques apportées par le NGS n'ont pas permis de résoudre encore à ce jour.

Afin de passer outre la difficulté d'interprétation de ces variations dans des gènes très éloignés des voies connues de la cancérogenèse, nous nous sommes appuyés sur le concept des panels de gènes en sélectionnant 201 gènes pour lesquels l'implication au niveau somatique ou constitutionnel dans la carcinogenèse était établie. Nous avons également maintenu le concept de phénotype extrême dans la sélection des cas analysés, afin de limiter le risque de faux-positifs et les difficultés d'interprétation en présence d'une population hétérogène et au risque génétique faible. En dépit d'un rationnel fort par la sélection d'une liste de gènes fréquemment altérés par des altérations *drivers* dans la cancérogenèse ou impliqués dans des formes mendéliennes de prédisposition aux cancers, **cette capture ciblée n'a cependant permis l'identification d'une variation causale de manière formelle que chez une seule patiente** atteinte de cancer du sein précoce (#24, variation délétère de *TP53* en mosaïque). Néanmoins, nous avons observé une tendance nette, bien que non significative, à un **enrichissement en variants rares affectant des gènes de la réparation de l'ADN** dans notre cohorte de cancers du sein précoce, et plus largement dans l'ensemble des cohortes soumises à cette capture ciblée. Certains de ces variants rares présentent en outre des caractéristiques permettant de suspecter une certaine pathogénicité, parfois également suggérée par la littérature. Nous observons enfin un certain nombre de variants rares, pathogènes à l'état récessif mais altérés sur un mode hétérozygote chez nos patientes, permettant d'alimenter l'**hypothèse d'un déterminisme oligogénique ou de susceptibilités environnementales accrues**. Il convient cependant de reconnaître qu'à la différence des pathologies neuro-psychiatriques (déficience

intellectuelle, troubles autistiques, maladie d'Alzheimer du sujet précoce), et de syndromes génétiques de présentations phénotypiques plus homogènes, **notre approche de déploiement des ressources du NGS dans la caractérisation moléculaire de la prédisposition héréditaire au cancer du sein n'apporte qu'une contribution limitée**. Ce constat est en adéquation avec l'analyse de la littérature, avec peu d'approches similaires rapportées laissant supposer des résultats également peu fructueux. La **problématique de la population cible** apparaît au travers de ces conclusions comme l'élément discutable, dans le sens où notre présupposé d'un probable déterminisme génétique pour des formes extrêmement précoces de cancer du sein peut être remis en question. Si la précocité du diagnostic demeure un argument en faveur d'une altération génétique constitutionnelle, et si la présentation sporadique amenait logiquement à suspecter des altérations *de novo*, l'absence de variation délétère détectée chez nos patientes doit également s'interpréter comme traduisant un **possible mécanisme accidentel somatique**. Il convient ici de rappeler l'étude de Tomasetti et Volgestein [330] introduisant une tierce composante dans la survenue des cancers : facteurs extrinsèques environnementaux, facteurs intrinsèques correspondant aux prédispositions héréditaires et risque d'acquisition somatique de variations tissu-dépendant. Ce dernier facteur, stochastique, est corrélé aux nombres de divisions des cellules souches au cours de la vie des individus, et ainsi plus élevé dans certains tissus que d'autres. Cette composante peut être aisément retenue dans le cancer du sein, la glande mammaire subissant au cours de la vie génitale de la femme des remaniements constants. Ainsi, si la survenue très précoce d'un cancer du sein appelle intuitivement à suspecter une altération génétique prédisposante, **il ne peut être exclu qu'elle ne résulte d'une altération somatique** liée à cette homéostasie particulièrement dynamique.

L'identification d'une variation délétère en mosaïque de *TP53* dans le contexte d'un cancer du sein bilatéral et précoce nous a enfin conduit à rechercher spécifiquement de tels événements dans des cohortes évocatrices d'un syndrome de Li-Fraumeni : cancers pédiatriques, tumeurs emblématiques du syndrome LFS dont le cancer du sein de la femme jeune, tumeurs primitives multiples. Nous avons pu ainsi rapporter une **fréquence importante de variations *de novo* de *TP53*, dont certaines en mosaïque**, contribuant ainsi à nuancer les critères cliniques justifiant une analyse *TP53*, notamment à ne pas exclure du champ d'investigation les formes sporadiques. Le séquençage de nouvelle génération devient également, au travers de cette étude, un outil incontournable pour l'analyse *TP53*.

Ainsi, au terme de ces différentes approches, nous concluons que l'apport du séquençage de nouvelle génération est indéniable dans l'exploration moléculaire des formes précoces de cancer, dans la recherche des possibles déterminants constitutionnels. Néanmoins, les approches conceptuelles développées ont montré des résultats nuancés et restreints, avec la difficulté, malgré des cohortes homogènes constituées sur le trait phénotypique de l'extrême précocité, à identifier au-delà de situations anecdotiques (variant *INHBA*, variations en mosaïque de *TP53*) de nouveaux *substratum* moléculaires. Ainsi, nous n'avons pu illustrer par ces séquençages exomiques ou par larges panels de gènes, le concept de variations privées largement suspecté dans ces formes sporadiques précoces. L'enrichissement en variants rares affectant les gènes

impliqués dans les voies de la réparation de l'ADN constitue néanmoins, bien que non significatif, une hypothèse de travail pour une exploration plus approfondie de ces voies, notamment du complexe de Fanconi. La problématique de l'oligogénisme, soit la contribution de plusieurs variants rares agissant dans une combinatoire défavorable, est pour l'heure difficilement appréciable. Dans ce modèle, le phénotype peut résulter d'un digénisme (deux altérations génétiques sont nécessaires pour provoquer le phénotype) voire d'un polygénisme, ou de la combinaison d'un monogénisme à faible pénétrance associé à un ou des facteurs modificateurs. Des bases de données récentes tentent de répertorier ces pathologies oligogéniques telles que ORVAL [331]. L'intelligence artificielle en développement pourrait constituer une aide à l'identification et à l'interprétation de ces variants agissant en oligogénisme [332, 333].

Enfin, l'identification de variations en mosaïque de *TP53* avec des prévalences significatives sur le plan clinique, doit conduire à des approches technologiques permettant d'appréhender ces événements, et conforte l'évolution des critères de Chompret conduisant à proposer une analyse *TP53* en **situation sporadique** pour les situations de cancers pédiatriques ou adultes rares tels que les corticosurrénalomes, les rhabdomyosarcomes embryonnaires de type anaplasique et les cancers du sein avant 31 ans, ou encore les tumeurs primitives multiples.

Bibliographie

1. Schwartz, A.G., et al., Risk of breast cancer to relatives of young breast cancer patients. *J Natl Cancer Inst*, 1985. **75**(4): p. 665-8.
2. Ottman, R., et al., Familial breast cancer in a population-based series. *Am J Epidemiol*, 1986. **123**(1): p. 15-21.
3. Claus, E.B., N.J. Risch, and W.D. Thompson, Age at onset as an indicator of familial risk of breast cancer. *Am J Epidemiol*, 1990. **131**(6): p. 961-72.
4. Newman, B., et al., Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sci U S A*, 1988. **85**(9): p. 3044-8.
5. Hall, J.M., et al., Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 1990. **250**(4988): p. 1684-9.
6. Hall, J.M., et al., Closing in on a breast cancer gene on chromosome 17q. *Am J Hum Genet*, 1992. **50**(6): p. 1235-42.
7. Narod, S.A., et al., Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet*, 1991. **338**(8759): p. 82-3.
8. Easton, D.F., et al., Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *Am J Hum Genet*, 1993. **52**(4): p. 678-701.
9. Miki, Y., et al., A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 1994. **266**(5182): p. 66-71.
10. Wooster, R., et al., Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*, 1994. **265**(5181): p. 2088-90.
11. Rass, E., et al., [Double strand break repair, one mechanism can hide another: alternative non-homologous end joining]. *Cancer Radiother*, 2012. **16**(1): p. 1-10.
12. Chock, K.L., et al., BRCA1-IRIS overexpression promotes cisplatin resistance in ovarian cancer cells. *Cancer Res*, 2010. **70**(21): p. 8782-91.
13. Shimizu, Y., et al., BRCA1-IRIS overexpression promotes formation of aggressive breast cancers. *PLoS One*, 2012. **7**(4): p. e34102.
14. Isono, M., et al., BRCA1 Directs the Repair Pathway to Homologous Recombination by Promoting 53BP1 Dephosphorylation. *Cell Rep*, 2017. **18**(2): p. 520-532.
15. Nepomuceno, T.C., et al., The Role of PALB2 in the DNA Damage Response and Cancer Predisposition. *Int J Mol Sci*, 2017. **18**(9).
16. Szabo, C., et al., The breast cancer information core: database design, structure, and scope. *Hum Mutat*, 2000. **16**(2): p. 123-31.
17. Fokkema, I.F., et al., LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*, 2011. **32**(5): p. 557-63.
18. Caputo, S., et al., Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. *Nucleic Acids Res*, 2012. **40**(Database issue): p. D992-1002.
19. Spurdle, A.B., et al., ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat*, 2012. **33**(1): p. 2-7.
20. Landrum, M.J., et al., ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*, 2014. **42**(Database issue): p. D980-5.
21. Cline, M.S., et al., BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet*, 2018. **14**(12): p. e1007752.
22. Richards, S., et al., 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. *Genet Med*, 2015. **17**(5): p. 405-24.
23. Starita, L.M., et al., Variant Interpretation: Functional Assays to the Rescue. *Am J Hum Genet*, 2017. **101**(3): p. 315-325.
24. Bonnet, C., et al., Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J Med Genet*, 2008. **45**(7): p. 438-46.

25. Thery, J.C., et al., Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet*, 2011. **19**(10): p. 1052-8.
26. Houdayer, C., et al., Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat*, 2012. **33**(8): p. 1228-38.
27. Gaildrat, P., et al., The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements. *J Med Genet*, 2010. **47**(6): p. 398-403.
28. Tammaro, C., et al., BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer. *Biochem Soc Trans*, 2012. **40**(4): p. 768-72.
29. Amendola, L.M., et al., Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*, 2016. **99**(1): p. 247.
30. Mersch, J., et al., Prevalence of Variant Reclassification Following Hereditary Cancer Genetic Testing. *JAMA*, 2018. **320**(12): p. 1266-1274.
31. Kuchenbaecker, K.B., et al., Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA*, 2017. **317**(23): p. 2402-2416.
32. Tai, Y.C., et al., Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst*, 2007. **99**(23): p. 1811-4.
33. Oh, M., et al., The association of BRCA1 and BRCA2 mutations with prostate cancer risk, frequency, and mortality: A meta-analysis. *Prostate*, 2019.
34. Iqbal, J., et al., The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers. *Br J Cancer*, 2012. **107**(12): p. 2005-9.
35. Ford, D., et al., Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Lancet*, 1994. **343**(8899): p. 692-5.
36. Breast Cancer Linkage, C., Cancer risks in BRCA2 mutation carriers. *J Natl Cancer Inst*, 1999. **91**(15): p. 1310-6.
37. Sharma, P., et al., Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing. *Breast Cancer Res Treat*, 2014. **145**(3): p. 707-14.
38. Hahnen, E., et al., Germline Mutation Status, Pathological Complete Response, and Disease-Free Survival in Triple-Negative Breast Cancer: Secondary Analysis of the GeparSixto Randomized Clinical Trial. *JAMA Oncol*, 2017. **3**(10): p. 1378-1385.
39. Consortium, B.C.L., Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium. *Lancet*, 1997. **349**(9064): p. 1505-10.
40. Spurdle, A.B., et al., Refined histopathological predictors of BRCA1 and BRCA2 mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res*, 2014. **16**(6): p. 3419.
41. Tun, N.M., et al., Risk of having BRCA1 mutation in high-risk women with triple-negative breast cancer: a meta-analysis. *Clin Genet*, 2014. **85**(1): p. 43-8.
42. Mavaddat, N., et al., Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer Epidemiol Biomarkers Prev*, 2012. **21**(1): p. 134-47.
43. Cancer, I.N.d. © /*Thésaurus - Femmes porteuses d'une mutation de BRCA1 ou BRCA2 / Détection précoce du cancer du sein et des annexes et stratégies de réduction du risque*. Collection recommandations et référentiels 2017; Available from: <https://www.e-cancer.fr/Professionnels-de-sante/Recommandations-et-outils-d-aide-a-la-pratique/Oncogenetique-ou-Cancers-avec-predispositions-genetiques>.
44. Kriege, M., et al., Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med*, 2004. **351**(5): p. 427-37.
45. Obdeijn, I.M., et al., Assessment of false-negative cases of breast MR imaging in women with a familial or genetic predisposition. *Breast Cancer Res Treat*, 2010. **119**(2): p. 399-407.
46. Rijnsburger, A.J., et al., BRCA1-associated breast cancers present differently from BRCA2-associated and familial cases: long-term follow-up of the Dutch MRISC Screening Study. *J Clin Oncol*, 2010. **28**(36): p. 5265-73.

47. Le-Petross, H.T., et al., Effectiveness of alternating mammography and magnetic resonance imaging for screening women with deleterious BRCA mutations at high risk of breast cancer. *Cancer*, 2011. **117**(17): p. 3900-7.
48. Riedl, C.C., et al., Triple-modality screening trial for familial breast cancer underlines the importance of magnetic resonance imaging and questions the role of mammography and ultrasound regardless of patient mutation status, age, and breast density. *J Clin Oncol*, 2015. **33**(10): p. 1128-35.
49. Moller, P., et al., Survival of patients with BRCA1-associated breast cancer diagnosed in an MRI-based surveillance program. *Breast Cancer Res Treat*, 2013. **139**(1): p. 155-61.
50. Evans, D.G., et al., Intensive breast screening in BRCA2 mutation carriers is associated with reduced breast cancer specific and all cause mortality. *Hered Cancer Clin Pract*, 2016. **14**: p. 8.
51. Sardanelli, F., et al., Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk italian 1 study): final results. *Invest Radiol*, 2011. **46**(2): p. 94-105.
52. Passaperuma, K., et al., Long-term results of screening with magnetic resonance imaging in women with BRCA mutations. *Br J Cancer*, 2012. **107**(1): p. 24-30.
53. Meijers-Heijboer, H., et al., Breast cancer after prophylactic bilateral mastectomy in women with a BRCA1 or BRCA2 mutation. *N Engl J Med*, 2001. **345**(3): p. 159-64.
54. Rebbeck, T.R., et al., Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *J Clin Oncol*, 2004. **22**(6): p. 1055-62.
55. Skytte, A.B., et al., Breast cancer after bilateral risk-reducing mastectomy. *Clin Genet*, 2011. **79**(5): p. 431-7.
56. Domchek, S.M., et al., Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA*, 2010. **304**(9): p. 967-75.
57. Heemskerk-Gerritsen, B.A., et al., Substantial breast cancer risk reduction and potential survival benefit after bilateral mastectomy when compared with surveillance in healthy BRCA1 and BRCA2 mutation carriers: a prospective analysis. *Ann Oncol*, 2013. **24**(8): p. 2029-35.
58. Giannakeas, V. and S.A. Narod, The expected benefit of preventive mastectomy on breast cancer incidence and mortality in BRCA mutation carriers, by age at mastectomy. *Breast Cancer Res Treat*, 2018. **167**(1): p. 263-267.
59. Kurian, A.W., B.M. Sigal, and S.K. Plevritis, Survival analysis of cancer risk reduction strategies for BRCA1/2 mutation carriers. *J Clin Oncol*, 2010. **28**(2): p. 222-31.
60. Sigal, B.M., et al., A simulation model to predict the impact of prophylactic surgery and screening on the life expectancy of BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev*, 2012. **21**(7): p. 1066-77.
61. Pierce, L.J., et al., Local therapy in BRCA1 and BRCA2 mutation carriers with operable breast cancer: comparison of breast conservation and mastectomy. *Breast Cancer Res Treat*, 2010. **121**(2): p. 389-98.
62. Valachis, A., A.D. Nearchou, and P. Lind, Surgical management of breast cancer in BRCA-mutation carriers: a systematic review and meta-analysis. *Breast Cancer Res Treat*, 2014. **144**(3): p. 443-55.
63. Evans, D.G., et al., Contralateral mastectomy improves survival in women with BRCA1/2-associated breast cancer. *Breast Cancer Res Treat*, 2013. **140**(1): p. 135-42.
64. Heemskerk-Gerritsen, B.A., et al., Improved overall survival after contralateral risk-reducing mastectomy in BRCA1/2 mutation carriers with a history of unilateral breast cancer: a prospective analysis. *Int J Cancer*, 2015. **136**(3): p. 668-77.
65. Marchetti, C., et al., Risk-reducing salpingo-oophorectomy: a meta-analysis on impact on ovarian cancer risk and all cause mortality in BRCA 1 and BRCA 2 mutation carriers. *BMC Womens Health*, 2014. **14**: p. 150.
66. Rebbeck, T.R., N.D. Kauff, and S.M. Domchek, Meta-analysis of risk reduction estimates associated with risk-reducing salpingo-oophorectomy in BRCA1 or BRCA2 mutation carriers. *J Natl Cancer Inst*, 2009. **101**(2): p. 80-7.
67. Kotsopoulos, J., et al., Oophorectomy after menopause and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev*, 2012. **21**(7): p. 1089-96.

68. Cuzick, J., et al., Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *Lancet*, 2013. **381**(9880): p. 1827-34.
69. Cuzick, J., et al., Anastrozole for prevention of breast cancer in high-risk postmenopausal women (IBIS-II): an international, double-blind, randomised placebo-controlled trial. *Lancet*, 2014. **383**(9922): p. 1041-8.
70. Group, T.A.B.C.A.S., Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. *Br J Cancer*, 2000. **83**(10): p. 1301-8.
71. Robson, M., et al., Prevalence of recurring BRCA mutations among Ashkenazi Jewish women with breast cancer. *Genet Test*, 1997. **1**(1): p. 47-51.
72. Górski, B., et al., A high proportion of founder BRCA1 mutations in Polish breast cancer families. *International Journal of Cancer*, 2004. **110**(5): p. 683-686.
73. Parmigiani, G., D. Berry, and O. Aguilar, Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet*, 1998. **62**(1): p. 145-58.
74. Lee, A.J., et al., BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer*, 2014. **110**(2): p. 535-45.
75. Evans, D.G., et al., A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. *J Med Genet*, 2004. **41**(6): p. 474-80.
76. Eisinger, F., et al., [Identification and management of hereditary predisposition to cancer of the breast and the ovary (update 2004)]. *Bull Cancer*, 2004. **91**(3): p. 219-37.
77. Statement of the American Society of Clinical Oncology: genetic testing for cancer susceptibility, Adopted on February 20, 1996. *J Clin Oncol*, 1996. **14**(5): p. 1730-6; discussion 1737-40.
78. Azzollini, J., et al., Mutation detection rates associated with specific selection criteria for BRCA1/2 testing in 1854 high-risk families: A monocentric Italian study. *Eur J Intern Med*, 2016. **32**: p. 65-71.
79. Collet, G., et al., Cost-effectiveness evaluation of pre-counseling telephone interviews before face-to-face genetic counseling in cancer genetics. *Fam Cancer*, 2018. **17**(3): p. 451-457.
80. Peto, J., et al., Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. *J Natl Cancer Inst*, 1999. **91**(11): p. 943-9.
81. Cui, J. and J.L. Hopper, Why are the majority of hereditary cases of early-onset breast cancer sporadic? A simulation study. *Cancer Epidemiol Biomarkers Prev*, 2000. **9**(8): p. 805-12.
82. Hartge, P., et al., The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am J Hum Genet*, 1999. **64**(4): p. 963-70.
83. Smith, E.D., B.K. Kennedy, and M. Kaerberlein, Genome-wide identification of conserved longevity genes in yeast and worms. *Mech Ageing Dev*, 2007. **128**(1): p. 106-11.
84. Kainu, T., et al., Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. *Proc Natl Acad Sci U S A*, 2000. **97**(17): p. 9603-8.
85. Thompson, D., et al., Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. *Proc Natl Acad Sci U S A*, 2002. **99**(2): p. 827-31.
86. Zuppan, P., et al., Possible linkage of the estrogen receptor gene to breast cancer in a family with late-onset disease. *Am J Hum Genet*, 1991. **48**(6): p. 1065-8.
87. Huusko, P., et al., Genome-wide scanning for linkage in Finnish breast cancer families. *Eur J Hum Genet*, 2004. **12**(2): p. 98-104.
88. Kerangueven, F., et al., Loss of heterozygosity and linkage analysis in breast carcinoma: indication for a putative third susceptibility gene on the short arm of chromosome 8. *Oncogene*, 1995. **10**(5): p. 1023-6.
89. Seitz, S., et al., Deletion mapping and linkage analysis provide strong indication for the involvement of the human chromosome region 8p12-p22 in breast carcinogenesis. *Br J Cancer*, 1997. **76**(8): p. 983-91.
90. Rahman, N., et al., Absence of evidence for a familial breast cancer susceptibility gene at chromosome 8p12-p22. *Oncogene*, 2000. **19**(36): p. 4170-3.
91. Gronwald, J., et al., Comparison of genomic abnormalities between BRCA1 and sporadic breast cancers studied by comparative genomic hybridization. *Int J Cancer*, 2005. **114**(2): p. 230-6.

92. Mangia, A., et al., Genetic heterogeneity by comparative genomic hybridization in BRCAx breast cancers. *Cancer Genet Cytogenet*, 2008. **182**(2): p. 75-83.
93. Didraga, M.A., et al., A non-BRCA1/2 hereditary breast cancer sub-group defined by aCGH profiling of genetically related patients. *Breast Cancer Res Treat*, 2011. **130**(2): p. 425-36.
94. Honrado, E., et al., Immunohistochemical classification of non-BRCA1/2 tumors identifies different groups that demonstrate the heterogeneity of BRCAx families. *Mod Pathol*, 2007. **20**(12): p. 1298-306.
95. Fernandez-Ramires, R., et al., Transcriptional characteristics of familial non-BRCA1/BRCA2 breast tumors. *Int J Cancer*, 2011. **128**(11): p. 2635-44.
96. Li, F.P., *Soft-Tissue Sarcomas, Breast Cancer, and Other Neoplasms*. *Annals of Internal Medicine*, 1969. **71**(4).
97. Li, F.P., et al., A cancer family syndrome in twenty-four kindreds. *Cancer Res*, 1988. **48**(18): p. 5358-62.
98. Birch, J.M., et al., Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families. *Cancer Res*, 1994. **54**(5): p. 1298-304.
99. Eeles, R.A., Germline mutations in the TP53 gene. *Cancer Surv*, 1995. **25**: p. 101-24.
100. Tinat, J., et al., 2009 version of the Chompret criteria for Li Fraumeni syndrome. *J Clin Oncol*, 2009. **27**(26): p. e108-9; author reply e110.
101. Bougeard, G., et al., Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *J Clin Oncol*, 2015. **33**(21): p. 2345-52.
102. Malkin, D., et al., Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*, 1990. **250**(4985): p. 1233-8.
103. Srivastava, S., et al., Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature*, 1990. **348**(6303): p. 747-9.
104. May, P. and E. May, Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene*, 1999. **18**(53): p. 7621-36.
105. Pluquet, O. and P. Hainaut, Genotoxic and non-genotoxic pathways of p53 induction. *Cancer Lett*, 2001. **174**(1): p. 1-15.
106. Ljungman, M., Dial 9-1-1 for p53: mechanisms of p53 activation by cellular stress. *Neoplasia*, 2000. **2**(3): p. 208-25.
107. Kanchi, K.L., et al., Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun*, 2014. **5**: p. 3156.
108. Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumours. *Nature*, 2012. **490**(7418): p. 61-70.
109. Olivier, M., et al., Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. *Cancer Res*, 2003. **63**(20): p. 6643-50.
110. Bougeard, G., et al., Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *J Med Genet*, 2008. **45**(8): p. 535-8.
111. Zerdoumi, Y., et al., Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. *Hum Mutat*, 2013. **34**(3): p. 453-61.
112. Laloo, F., et al., Prediction of pathogenic mutations in patients with early-onset breast cancer by family history. *Lancet*, 2003. **361**(9363): p. 1101-2.
113. Walsh, T., et al., Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA*, 2006. **295**(12): p. 1379-88.
114. Ginsburg, O.M., et al., The prevalence of germ-line TP53 mutations in women diagnosed with breast cancer before age 30. *Fam Cancer*, 2009. **8**(4): p. 563-7.
115. Gonzalez, K.D., et al., High frequency of de novo mutations in Li-Fraumeni syndrome. *J Med Genet*, 2009. **46**(10): p. 689-93.
116. Mouchawar, J., et al., Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res*, 2010. **70**(12): p. 4795-800.
117. Ruijs, M.W., et al., TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. *J Med Genet*, 2010. **47**(6): p. 421-8.
118. McCuaig, J.M., et al., Routine TP53 testing for breast cancer under age 30: ready for prime time? *Fam Cancer*, 2012. **11**(4): p. 607-13.

119. Castera, L., et al., Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet*, 2014. **22**(11): p. 1305-13.
120. Bakhuizen, J.J., et al., TP53 germline mutation testing in early-onset breast cancer: findings from a nationwide cohort. *Fam Cancer*, 2019. **18**(2): p. 273-280.
121. Masciari, S., et al., Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. *Breast Cancer Res Treat*, 2012. **133**(3): p. 1125-30.
122. Melhem-Bertrandt, A., et al., Early onset HER2-positive breast cancer is associated with germline TP53 mutations. *Cancer*, 2012. **118**(4): p. 908-13.
123. Rath, M.G., et al., Prevalence of germline TP53 mutations in HER2+ breast cancer patients. *Breast Cancer Res Treat*, 2013. **139**(1): p. 193-8.
124. Evans, D.G., 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 Cancer*, 2010. **102**(7): p. 1091-8.
125. Villani, A., et al., Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. *Lancet Oncol*, 2011. **12**(6): p. 559-67.
126. Villani, A., et al., Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. *Lancet Oncol*, 2016. **17**(9): p. 1295-305.
127. Kratz, C.P., et al., Cancer Screening Recommendations for Individuals with Li-Fraumeni Syndrome. *Clin Cancer Res*, 2017. **23**(11): p. e38-e45.
128. Kasper, E., et al., Contribution of genotoxic anticancer treatments to the development of multiple primary tumours in the context of germline TP53 mutations. *Eur J Cancer*, 2018. **101**: p. 254-262.
129. Rahman, N., et al., PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*, 2007. **39**(2): p. 165-7.
130. Tischkowitz, M., et al., Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci U S A*, 2007. **104**(16): p. 6788-93.
131. Janatova, M., et al., The PALB2 gene is a strong candidate for clinical testing in BRCA1- and BRCA2-negative hereditary breast cancer. *Cancer Epidemiol Biomarkers Prev*, 2013. **22**(12): p. 2323-32.
132. Fernandes, P.H., et al., Comprehensive sequencing of PALB2 in patients with breast cancer suggests PALB2 mutations explain a subset of hereditary breast cancer. *Cancer*, 2014. **120**(7): p. 963-7.
133. Erkkö, H., et al., A recurrent mutation in PALB2 in Finnish cancer families. *Nature*, 2007. **446**(7133): p. 316-9.
134. Antoniou, A.C., et al., Breast-cancer risk in families with mutations in PALB2. *N Engl J Med*, 2014. **371**(6): p. 497-506.
135. Meindl, A., et al., Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet*, 2010. **42**(5): p. 410-4.
136. Loveday, C., et al., Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet*, 2011. **43**(9): p. 879-882.
137. Golmard, L., et al., Contribution of germline deleterious variants in the RAD51 paralogs to breast and ovarian cancers. *Eur J Hum Genet*, 2017. **25**(12): p. 1345-1353.
138. Song, H., et al., Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. *J Clin Oncol*, 2015. **33**(26): p. 2901-7.
139. Swisher, E.M., et al., Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol*, 2017. **18**(1): p. 75-87.
140. Damiola, F., et al., Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res*, 2014. **16**(3): p. R58.
141. Pilarski, R., et al., Cowden syndrome and the PTEN hamartoma tumor syndrome: systematic review and revised diagnostic criteria. *J Natl Cancer Inst*, 2013. **105**(21): p. 1607-16.
142. Pilarski, R., et al., Predicting PTEN mutations: an evaluation of Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome clinical features. *J Med Genet*, 2011. **48**(8): p. 505-12.

143. Tan, M.H., et al., A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *Am J Hum Genet*, 2011. **88**(1): p. 42-56.
144. Tan, M.H., et al., Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res*, 2012. **18**(2): p. 400-7.
145. Bubien, V., et al., High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. *J Med Genet*, 2013. **50**(4): p. 255-63.
146. Nieuwenhuis, M.H., et al., Cancer risk and genotype-phenotype correlations in PTEN hamartoma tumor syndrome. *Fam Cancer*, 2014. **13**(1): p. 57-63.
147. Ni, Y., et al., Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. *Hum Mol Genet*, 2012. **21**(2): p. 300-10.
148. Hobert, J.A. and C. Eng, PTEN hamartoma tumor syndrome: an overview. *Genet Med*, 2009. **11**(10): p. 687-94.
149. Lim, W., et al., Further observations on LKB1/STK11 status and cancer risk in Peutz-Jeghers syndrome. *Br J Cancer*, 2003. **89**(2): p. 308-13.
150. Pharoah, P.D., et al., Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*, 2001. **121**(6): p. 1348-53.
151. Hansford, S., et al., Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol*, 2015. **1**(1): p. 23-32.
152. Corso, G., et al., Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. *J Med Genet*, 2018. **55**(7): p. 431-441.
153. van der Post, R.S., et al., Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet*, 2015. **52**(6): p. 361-74.
154. Corso, G., et al., CDH1 germline mutations and hereditary lobular breast cancer. *Fam Cancer*, 2016. **15**(2): p. 215-9.
155. van Marcke, C., et al., Prevalence of pathogenic variants and variants of unknown significance in patients at high risk of breast cancer: A systematic review and meta-analysis of gene-panel data. *Crit Rev Oncol Hematol*, 2018. **132**: p. 138-144.
156. Thompson, D., et al., Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst*, 2005. **97**(11): p. 813-22.
157. Renwick, A., et al., ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet*, 2006. **38**(8): p. 873-5.
158. Gatti, R.A., A. Tward, and P. Concannon, Cancer risk in ATM heterozygotes: a model of phenotypic and mechanistic differences between missense and truncating mutations. *Mol Genet Metab*, 1999. **68**(4): p. 419-23.
159. Sodha, N., et al., Screening hCHK2 for mutations. *Science*, 2000. **289**(5478): p. 359.
160. Meijers-Heijboer, H., et al., Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet*, 2002. **31**(1): p. 55-9.
161. Vahteristo, P., et al., A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet*, 2002. **71**(2): p. 432-8.
162. Oldenburg, R.A., et al., The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. *Cancer Res*, 2003. **63**(23): p. 8153-7.
163. Seal, S., et al., Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*, 2006. **38**(11): p. 1239-41.
164. Easton, D.F., et al., No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk: implications for gene panel testing. *J Med Genet*, 2016. **53**(5): p. 298-309.
165. Buys, S.S., et al., A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer*, 2017. **123**(10): p. 1721-1730.
166. Weber-Lassalle, N., et al., BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. *Breast Cancer Res*, 2018. **20**(1): p. 7.
167. Ghimenti, C., et al., Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes Chromosomes Cancer*, 2002. **33**(3): p. 235-42.
168. Karppinen, S.M., et al., Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. *J Med Genet*, 2004. **41**(9): p. e114.

169. Suszynska, M., et al., BARD1 is A Low/Moderate Breast Cancer Risk Gene: Evidence Based on An Association Study of the Central European p.Q564X Recurrent Mutation. *Cancers* (Basel), 2019. **11**(6).
170. Spurdle, A.B., et al., Common genetic variation at BARD1 is not associated with breast cancer risk in BRCA1 or BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev*, 2011. **20**(5): p. 1032-8.
171. Weber-Lassalle, N., et al., Germline loss-of-function variants in the BARD1 gene are associated with early-onset familial breast cancer but not ovarian cancer. *Breast Cancer Res*, 2019. **21**(1): p. 55.
172. Shimelis, H., et al., Triple-Negative Breast Cancer Risk Genes Identified by Multigene Hereditary Cancer Panel Testing. *J Natl Cancer Inst*, 2018.
173. Park, D.J., et al., Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov*, 2014. **4**(7): p. 804-15.
174. Li, N., et al., Reevaluation of RINT1 as a breast cancer predisposition gene. *Breast Cancer Res Treat*, 2016. **159**(2): p. 385-92.
175. Peterlongo, P., et al., FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet*, 2015. **24**(18): p. 5345-55.
176. Landwehr, R., et al., Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer. *Breast Cancer Res Treat*, 2011. **130**(3): p. 1021-8.
177. Catucci, I., et al., Sequencing analysis of SLX4/FANCP gene in Italian familial breast cancer cases. *PLoS One*, 2012. **7**(2): p. e31038.
178. Fernandez-Rodriguez, J., et al., Analysis of SLX4/FANCP in non-BRCA1/2-mutated breast cancer families. *BMC Cancer*, 2012. **12**: p. 84.
179. Shah, S., et al., Assessment of SLX4 Mutations in Hereditary Breast Cancers. *PLoS One*, 2013. **8**(6): p. e66961.
180. Cybulski, C., et al., Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet*, 2015. **47**(6): p. 643-6.
181. Sun, J., et al., Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. *PLoS Genet*, 2015. **11**(5): p. e1005228.
182. Kwong, A., et al., Germline RECQL mutations in high risk Chinese breast cancer patients. *Breast Cancer Res Treat*, 2016. **157**(2): p. 211-215.
183. Bogdanova, N., et al., Analysis of a RECQL splicing mutation, c.1667_1667+3delAGTA, in breast cancer patients and controls from Central Europe. *Fam Cancer*, 2017. **16**(2): p. 181-186.
184. Li, N., et al., Mutations in RECQL are not associated with breast cancer risk in an Australian population. *Nat Genet*, 2018. **50**(10): p. 1346-1348.
185. Nguyen-Dumont, T., et al., FANCM and RECQL genetic variants and breast cancer susceptibility: relevance to South Poland and West Ukraine. *BMC Med Genet*, 2018. **19**(1): p. 12.
186. Tervasmaki, A., et al., Rare missense mutations in RECQL and POLG associate with inherited predisposition to breast cancer. *Int J Cancer*, 2018. **142**(11): p. 2286-2292.
187. Watson, P. and H.T. Lynch, Extracolonic cancer in hereditary nonpolyposis colorectal cancer. *Cancer*, 1993. **71**(3): p. 677-85.
188. Helder-Woolderink, J.M., et al., Ovarian cancer in Lynch syndrome; a systematic review. *Eur J Cancer*, 2016. **55**: p. 65-73.
189. Bonadona, V., et al., Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA*, 2011. **305**(22): p. 2304-10.
190. Bennett, J.A., et al., Mismatch Repair Protein Expression in Clear Cell Carcinoma of the Ovary: Incidence and Morphologic Associations in 109 Cases. *Am J Surg Pathol*, 2016. **40**(5): p. 656-63.
191. Bennett, J.A., et al., Incidence of Mismatch Repair Protein Deficiency and Associated Clinicopathologic Features in a Cohort of 104 Ovarian Endometrioid Carcinomas. *Am J Surg Pathol*, 2019. **43**(2): p. 235-243.
192. Le, D.T., et al., PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*, 2015. **372**(26): p. 2509-20.
193. Goldberg, M., et al., Association between the Lynch syndrome gene MSH2 and breast cancer susceptibility in a Canadian familial cancer registry. *J Med Genet*, 2017. **54**(11): p. 742-746.

194. Win, A.K., et al., Risks of colorectal and other cancers after endometrial cancer for women with Lynch syndrome. *J Natl Cancer Inst*, 2013. **105**(4): p. 274-9.
195. Buerki, N., et al., Evidence for breast cancer as an integral part of Lynch syndrome. *Genes Chromosomes Cancer*, 2012. **51**(1): p. 83-91.
196. Clay, M.R., et al., Risk of secondary malignancy (including breast) in patients with mismatch-repair protein deficiency. *Am J Surg Pathol*, 2014. **38**(11): p. 1494-500.
197. Wasielewski, M., et al., Association of rare MSH6 variants with familial breast cancer. *Breast Cancer Res Treat*, 2010. **123**(2): p. 315-20.
198. Win, A.K., N.M. Lindor, and M.A. Jenkins, Risk of breast cancer in Lynch syndrome: a systematic review. *Breast Cancer Res*, 2013. **15**(2): p. R27.
199. Castera, L., et al., Landscape of pathogenic variations in a panel of 34 genes and cancer risk estimation from 5131 HBOC families. *Genet Med*, 2018.
200. Kohane, I.S., M. Hsing, and S.W. Kong, Taxonomizing, sizing, and overcoming the incidentalome. *Genet Med*, 2012. **14**(4): p. 399-404.
201. Balmana, J., et al., Conflicting Interpretation of Genetic Variants and Cancer Risk by Commercial Laboratories as Assessed by the Prospective Registry of Multiplex Testing. *J Clin Oncol*, 2016. **34**(34): p. 4071-4078.
202. Plon, S.E., et al., Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*, 2008. **29**(11): p. 1282-91.
203. Parsons, M.T., et al., Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Hum Mutat*, 2019.
204. Thompson, B.A., et al., Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*, 2014. **46**(2): p. 107-115.
205. Kurian, A.W., et al., Gaps in Incorporating Germline Genetic Testing Into Treatment Decision-Making for Early-Stage Breast Cancer. *J Clin Oncol*, 2017. **35**(20): p. 2232-2239.
206. Young, E.L., et al., Multigene testing of moderate-risk genes: be mindful of the missense. *J Med Genet*, 2016. **53**(6): p. 366-76.
207. Thompson, E.R., et al., Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. *J Clin Oncol*, 2016. **34**(13): p. 1455-9.
208. Song, W., et al., Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. *Genet Med*, 2016. **18**(8): p. 850-4.
209. NCBI. *Genetic Testing Registry*. 14 Mai 2019]; Available from: <https://www-ncbi-nlm-nih-gov.gate2.inist.fr/gtr/>.
210. Couch, F.J., et al., Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol*, 2015. **33**(4): p. 304-11.
211. Ellingson, M.S., et al., Exome sequencing reveals frequent deleterious germline variants in cancer susceptibility genes in women with invasive breast cancer undergoing neoadjuvant chemotherapy. *Breast Cancer Res Treat*, 2015. **153**(2): p. 435-43.
212. Li, J., et al., Targeted massively parallel sequencing of a panel of putative breast cancer susceptibility genes in a large cohort of multiple-case breast and ovarian cancer families. *J Med Genet*, 2016. **53**(1): p. 34-42.
213. Couch, F.J., et al., Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol*, 2017. **3**(9): p. 1190-1196.
214. Kraus, C., et al., Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int J Cancer*, 2017. **140**(1): p. 95-102.
215. Hauke, J., et al., Gene panel testing of 5589 BRCA1/2-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer. *Cancer Med*, 2018. **7**(4): p. 1349-1358.
216. Girard, E., et al., Familial breast cancer and DNA repair genes: Insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing. *Int J Cancer*, 2019. **144**(8): p. 1962-1974.

217. Melchor, L. and J. Benitez, The complex genetic landscape of familial breast cancer. *Hum Genet*, 2013. **132**(8): p. 845-63.
218. Anczukow, O., et al., BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy. *Clin Cancer Res*, 2012. **18**(18): p. 4903-9.
219. Byers, H., et al., Sensitivity of BRCA1/2 testing in high-risk breast/ovarian/male breast cancer families: little contribution of comprehensive RNA/NGS panel testing. *Eur J Hum Genet*, 2016. **24**(11): p. 1591-1597.
220. Evans, D.G.R., et al., A Dominantly Inherited 5' UTR Variant Causing Methylation-Associated Silencing of BRCA1 as a Cause of Breast and Ovarian Cancer. *Am J Hum Genet*, 2018. **103**(2): p. 213-220.
221. Lheureux, S., et al., Two novel variants in the 3'UTR of the BRCA1 gene in familial breast and/or ovarian cancer. *Breast Cancer Res Treat*, 2011. **125**(3): p. 885-91.
222. Brewster, B.L., et al., Identification of fifteen novel germline variants in the BRCA1 3'UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. *Hum Mutat*, 2012. **33**(12): p. 1665-75.
223. de Mayo, T., et al., Identification of a Rare Germline Heterozygous Deletion Involving the Polycistronic miR-17-92 Cluster in Two First-Degree Relatives from a BRCA 1/2 Negative Chilean Family with Familial Breast Cancer: Possible Functional Implications. *Int J Mol Sci*, 2018. **19**(1).
224. Michailidou, K., et al., Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*, 2013. **45**(4): p. 353-61, 361e1-2.
225. International Schizophrenia, C., et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 2009. **460**(7256): p. 748-52.
226. Cross-Disorder Group of the Psychiatric Genomics, C., et al., Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*, 2013. **45**(9): p. 984-94.
227. Mavaddat, N., et al., Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*, 2019. **104**(1): p. 21-34.
228. Frebourg, T., The challenge for the next generation of medical geneticists. *Hum Mutat*, 2014. **35**(8): p. 909-11.
229. Wen, H., et al., Family-specific, novel, deleterious germline variants provide a rich resource to identify genetic predispositions for BRCAx familial breast cancer. *BMC Cancer*, 2014. **14**: p. 470.
230. Monies, D., et al., The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum Genet*, 2017. **136**(8): p. 921-939.
231. Conrad, D.F., et al., Variation in genome-wide mutation rates within and between human families. *Nat Genet*, 2011. **43**(7): p. 712-4.
232. Veltman, J.A. and H.G. Brunner, De novo mutations in human genetic disease. *Nat Rev Genet*, 2012. **13**(8): p. 565-75.
233. Chan, K. and D.A. Gordenin, Clusters of Multiple Mutations: Incidence and Molecular Mechanisms. *Annu Rev Genet*, 2015. **49**: p. 243-67.
234. Goriely, A. and A.O. Wilkie, Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet*, 2012. **90**(2): p. 175-200.
235. Acuna-Hidalgo, R., J.A. Veltman, and A. Hoischen, New insights into the generation and role of de novo mutations in health and disease. *Genome Biol*, 2016. **17**(1): p. 241.
236. Lynch, M., Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*, 2010. **107**(3): p. 961-8.
237. Vissers, L.E., et al., A de novo paradigm for mental retardation. *Nat Genet*, 2010. **42**(12): p. 1109-12.
238. Nicolas, G. and J.A. Veltman, The role of de novo mutations in adult-onset neurodegenerative disorders. *Acta Neuropathol*, 2019. **137**(2): p. 183-207.
239. Boycott, K.M., et al., Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 2013. **14**(10): p. 681-91.
240. Yang, Y., et al., Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*, 2014. **312**(18): p. 1870-9.
241. Posey, J.E., et al., Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med*, 2016. **18**(7): p. 678-85.

242. Lelieveld, S.H., et al., Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci*, 2016. **19**(9): p. 1194-6.
243. Wang, W., R. Corominas, and G.N. Lin, De novo Mutations From Whole Exome Sequencing in Neurodevelopmental and Psychiatric Disorders: From Discovery to Application. *Front Genet*, 2019. **10**: p. 258.
244. Imamura, H., et al., Sporadic paraganglioma caused by de novo SDHB mutations in a 6-year-old girl. *Eur J Pediatr*, 2016. **175**(1): p. 137-41.
245. Ng, S.B., et al., Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*, 2010. **42**(9): p. 790-3.
246. Petrovski, S., et al., Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*, 2013. **9**(8): p. e1003709.
247. Samocha, K.E., et al., A framework for the interpretation of de novo mutation in human disease. *Nat Genet*, 2014. **46**(9): p. 944-50.
248. Lek, M., et al., Analysis of protein-coding genetic variation in 60,706 humans. 2016. **536**(7616): p. 285-91.
249. Konrad J Karczewski, L.C.F., Grace Tiao, Beryl B Cummings, Jessica Alföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel P Birnbaum, Laura D Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A Watts, Daniel Rhodes, Moriel Singer-Berk, Eleanor G Seaby, Jack A Kosmicki, Raymond K Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X Chong, Kaitlin E Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S Ware, Christopher Vittal, Irina M Armean, Louis Bergelson, Kristian Cibulskis, Kristen M Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferreira, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E Talkowski, The Genome Aggregation Database Consortium, Benjamin M Neale, Mark J Daly, Daniel G MacArthur, *Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes*. *BioRxiv*, 2019.
250. Auer, P.L. and G. Lettre, Rare variant association studies: considerations, challenges and opportunities. *Genome Med*, 2015. **7**(1): p. 16.
251. Eilbeck, K., A. Quinlan, and M. Yandell, Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet*, 2017. **18**(10): p. 599-612.
252. Tewhey, R., et al., Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol*, 2009. **10**(10): p. R116.
253. Coutant, S., et al., EVA: Exome Variation Analyzer, an efficient and versatile tool for filtering strategies in medical genomics. *BMC Bioinformatics*, 2012. **13 Suppl 14**: p. S9.
254. Knight, P.G., L. Satchell, and C. Glister, *Intra-ovarian roles of activins and inhibins*. *Mol Cell Endocrinol*, 2012. **359**(1-2): p. 53-65.
255. Walton, K.L., Y. Makanji, and C.A. Harrison, New insights into the mechanisms of activin action and inhibition. *Mol Cell Endocrinol*, 2012. **359**(1-2): p. 2-12.
256. Matzuk, M.M., et al., Alpha-inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature*, 1992. **360**(6402): p. 313-9.
257. Di Simone, N., et al., Characterization of inhibin/activin subunit, follistatin, and activin type II receptors in human ovarian cancer cell lines: a potential role in autocrine growth regulation. *Endocrinology*, 1996. **137**(2): p. 486-94.
258. Morgenthaler, S. and W.G. Thilly, A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res*, 2007. **615**(1-2): p. 28-56.
259. Li, B. and S.M. Leal, Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*, 2008. **83**(3): p. 311-21.
260. Neale, B.M., et al., Testing for an unusual distribution of rare variants. *PLoS Genet*, 2011. **7**(3): p. e1001322.
261. Wu, M.C., et al., Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*, 2011. **89**(1): p. 82-93.
262. Mankanji, Y., et al., Suppression of inhibin A biological activity by alterations in the binding site for betaglycan. *J Biol Chem*, 2008. **283**(24): p. 16743-51.

263. Armstrong, C.T., J.L. Anderson, and R.M. Denton, Studies on the regulation of the human E1 subunit of the 2-oxoglutarate dehydrogenase complex, including the identification of a novel calcium-binding site. *Biochem J*, 2014. **459**(2): p. 369-81.
264. Dhanao, B.S., et al., Update on the Kelch-like (KLHL) gene family. *Hum Genomics*, 2013. **7**: p. 13.
265. Ng, S.B., et al., Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 2009. **461**(7261): p. 272-6.
266. Dorschner, M.O., et al., Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet*, 2013. **93**(4): p. 631-40.
267. Lawrence, M.S., et al., Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 2014. **505**(7484): p. 495-501.
268. Vogelstein, B., et al., *Cancer genome landscapes*. *Science*, 2013. **339**(6127): p. 1546-58.
269. Alexandrov, L.B., et al., Signatures of mutational processes in human cancer. *Nature*, 2013. **500**(7463): p. 415-21.
270. Samstein, R.M., et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*, 2019. **51**(2): p. 202-206.
271. Jhian, S.M., The RET proto-oncogene in human cancers. *Oncogene*, 2000. **19**(49): p. 5590-7.
272. Kim, H., et al., The association between non-breast and ovary cancers and BRCA mutation in first- and second-degree relatives of high-risk breast cancer patients: a large-scale study of Koreans. *Hered Cancer Clin Pract*, 2019. **17**: p. 1.
273. Prochazkova, K., et al., Somatic TP53 mutation mosaicism in a patient with Li-Fraumeni syndrome. *Am J Med Genet A*, 2009. **149A**(2): p. 206-11.
274. Behjati, S., et al., A pathogenic mosaic TP53 mutation in two germ layers detected by next generation sequencing. *PLoS One*, 2014. **9**(5): p. e96531.
275. Friedman, E., et al., Low-level constitutional mosaicism of a de novo BRCA1 gene mutation. *Br J Cancer*, 2015. **112**(4): p. 765-8.
276. Perou, C.M., et al., Molecular portraits of human breast tumours. *Nature*, 2000. **406**(6797): p. 747-52.
277. Nik-Zainal, S., et al., Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 2016. **534**(7605): p. 47-54.
278. Santarpia, L., et al., Deciphering and Targeting Oncogenic Mutations and Pathways in Breast Cancer. *Oncologist*, 2016. **21**(9): p. 1063-78.
279. Koboldt, D.C., et al., VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*, 2012. **22**(3): p. 568-76.
280. Backenroth, D., et al., CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res*, 2014. **42**(12): p. e97.
281. Ben Rekaya, M., et al., High frequency of the V548A fs X572 XPC mutation in Tunisia: implication for molecular diagnosis. *J Hum Genet*, 2009. **54**(7): p. 426-9.
282. He, J., et al., Associations of Lys939Gln and Ala499Val polymorphisms of the XPC gene with cancer susceptibility: a meta-analysis. *Int J Cancer*, 2013. **133**(8): p. 1765-75.
283. Bai, X., et al., Clinicopathological significance and prognostic value of Xeroderma pigmentosum complementary group C (XPC) expression in sporadic breast cancer patients. *Med Oncol*, 2012. **29**(3): p. 1543-53.
284. Thompson, E.R., et al., Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. *Hum Mutat*, 2012. **33**(1): p. 95-9.
285. Jonson, L., et al., Identification of six pathogenic RAD51C mutations via mutational screening of 1228 Danish individuals with increased risk of hereditary breast and/or ovarian cancer. *Breast Cancer Res Treat*, 2016. **155**(2): p. 215-22.
286. Yablonski-Peretz, T., et al., Screening for germline mutations in breast/ovarian cancer susceptibility genes in high-risk families in Israel. *Breast Cancer Res Treat*, 2016. **155**(1): p. 133-8.
287. Fulk, K., et al., Monoallelic MUTYH carrier status is not associated with increased breast cancer risk in a multigene panel cohort. *Fam Cancer*, 2019. **18**(2): p. 197-201.
288. Rennert, G., et al., MutYH mutation carriers have increased breast cancer risk. *Cancer*, 2012. **118**(8): p. 1989-93.
289. Win, A.K., et al., Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. *Int J Cancer*, 2016. **139**(7): p. 1557-63.

290. Vijai, J., et al., A Recurrent ERCC3 Truncating Mutation Confers Moderate Risk for Breast Cancer. *Cancer Discov*, 2016. **6**(11): p. 1267-1275.
291. Tournier, I., et al., A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat*, 2008. **29**(12): p. 1412-24.
292. Zhang, X., et al., Promoter hypermethylation of ARID1A gene is responsible for its low mRNA expression in many invasive breast cancers. *PLoS One*, 2013. **8**(1): p. e53931.
293. Zhao, J., C. Liu, and Z. Zhao, ARID1A: a potential prognostic factor for breast cancer. *Tumour Biol*, 2014. **35**(5): p. 4813-9.
294. Peron, A., et al., Do patients with tuberous sclerosis complex have an increased risk for malignancies? *Am J Med Genet A*, 2016. **170**(6): p. 1538-44.
295. Rump, A., et al., Identification and Functional Testing of ERCC2 Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer. *PLoS Genet*, 2016. **12**(8): p. e1006248.
296. Kattan, Z., et al., Damaged DNA binding protein 2 plays a role in breast cancer cell growth. *PLoS One*, 2008. **3**(4): p. e2002.
297. Sokolenko, A.P., et al., Candidate gene analysis of BRCA1/2 mutation-negative high-risk Russian breast cancer patients. *Cancer Lett*, 2015. **359**(2): p. 259-61.
298. Marutani, M., et al., Dominant-negative mutations of the tumor suppressor p53 relating to early onset of glioblastoma multiforme. *Cancer Res*, 1999. **59**(19): p. 4765-9.
299. Hassan, N.M., et al., Presence of dominant negative mutation of TP53 is a risk of early recurrence in oral cancer. *Cancer Lett*, 2008. **270**(1): p. 108-19.
300. Singh, T.R., et al., Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood*, 2009. **114**(1): p. 174-80.
301. Silvestri, V., et al., A possible role of FANCM mutations in male breast cancer susceptibility: Results from a multicenter study in Italy. *Breast*, 2018. **38**: p. 92-97.
302. Kiiski, J.I., et al., Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A*, 2014. **111**(42): p. 15172-7.
303. Hahnen, E., et al., Germline Mutations in Triple-Negative Breast Cancer. *Breast Care (Basel)*, 2017. **12**(1): p. 15-19.
304. Chandrasekharappa, S.C., et al., Assessing the spectrum of germline variation in Fanconi anemia genes among patients with head and neck carcinoma before age 50. *Cancer*, 2017. **123**(20): p. 3943-3954.
305. Degrolard-Courcet, E., et al., Development of primary early-onset colorectal cancers due to biallelic mutations of the FANCD1/BRCA2 gene. *Eur J Hum Genet*, 2014. **22**(8): p. 979-87.
306. Byrd, P.J., et al., A Hypomorphic PALB2 Allele Gives Rise to an Unusual Form of FA-N Associated with Lymphoid Tumour Development. *PLoS Genet*, 2016. **12**(3): p. e1005945.
307. Abbasi, S. and M. Rasouli, A rare FANCA gene variation as a breast cancer susceptibility allele in an Iranian population. *Mol Med Rep*, 2017. **15**(6): p. 3983-3988.
308. Mantere, T., et al., Case-control analysis of truncating mutations in DNA damage response genes connects TEX15 and FANCD2 with hereditary breast cancer susceptibility. *Sci Rep*, 2017. **7**(1): p. 681.
309. Penkert, J., et al., Breast cancer patients suggestive of Li-Fraumeni syndrome: mutational spectrum, candidate genes, and unexplained heredity. *Breast Cancer Res*, 2018. **20**(1): p. 87.
310. Kiiski, J.I., et al., FANCM mutation c.5791C>T is a risk factor for triple-negative breast cancer in the Finnish population. *Breast Cancer Res Treat*, 2017. **166**(1): p. 217-226.
311. Neidhardt, G., et al., Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. *JAMA Oncol*, 2017. **3**(9): p. 1245-1248.
312. Maung, K.Z.Y., et al., Rare variants in Fanconi anemia genes are enriched in acute myeloid leukemia. *Blood Cancer J*, 2018. **8**(6): p. 50.
313. Berwick, M., et al., Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. *Cancer Res*, 2007. **67**(19): p. 9591-6.
314. de Voer, R.M., et al., Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. *Sci Rep*, 2015. **5**: p. 14060.
315. Cleary, S.P., et al., Heterozygosity for the BLM(Ash) mutation and cancer risk. *Cancer Res*, 2003. **63**(8): p. 1769-71.
316. di Masi, A. and A. Antoccia, NBS1 Heterozygosity and Cancer Risk. *Curr Genomics*, 2008. **9**(4): p. 275-81.

317. Dumon-Jones, V., et al., Nbn heterozygosity renders mice susceptible to tumor formation and ionizing radiation-induced tumorigenesis. *Cancer Res*, 2003. **63**(21): p. 7263-9.
318. Lu, S., et al., Atm-haploinsufficiency enhances susceptibility to carcinogen-induced mammary tumors. *Carcinogenesis*, 2006. **27**(4): p. 848-55.
319. Ceccaldi, R., P. Sarangi, and A.D. D'Andrea, The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*, 2016. **17**(6): p. 337-49.
320. Chan, S.H., et al., Germline Mutations in Cancer Predisposition Genes are Frequent in Sporadic Sarcomas. *Sci Rep*, 2017. **7**(1): p. 10660.
321. Ballinger, M.L., et al., Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol*, 2016. **17**(9): p. 1261-71.
322. Gilissen, C., et al., Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 2014. **511**(7509): p. 344-7.
323. Zhang, J., et al., Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*, 2015. **373**(24): p. 2336-2346.
324. Charbonnier, F., et al., Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments. *Cancer Res*, 2000. **60**(11): p. 2760-3.
325. Bougeard, G., et al., Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. *Oncogene*, 2003. **22**(6): p. 840-6.
326. Baert-Desurmont, S., et al., Clinical relevance of 8q23, 15q13 and 18q21 SNP genotyping to evaluate colorectal cancer risk. *Eur J Hum Genet*, 2016. **24**(1): p. 99-105.
327. Swisher, E.M., et al., Somatic Mosaic Mutations in PPM1D and TP53 in the Blood of Women With Ovarian Carcinoma. *JAMA Oncol*, 2016. **2**(3): p. 370-2.
328. Coffee, B., et al., A Substantial Proportion of Apparently Heterozygous TP53 Pathogenic Variants Detected with a Next-Generation Sequencing Hereditary Pan-Cancer Panel Are Acquired Somatic. *Hum Mutat*, 2019.
329. Spier, I., et al., Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. *J Med Genet*, 2016. **53**(3): p. 172-9.
330. Tomasetti, C. and B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, 2015. **347**(6217): p. 78-81.
331. Renaux, A., et al., ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Res*, 2019. **47**(W1): p. W93-W98.
332. Papadimitriou, S., et al., Predicting disease-causing variant combinations. *Proc Natl Acad Sci U S A*, 2019. **116**(24): p. 11878-11887.
333. Versbaegen, N., et al., Using game theory and decision decomposition to effectively discern and characterise bi-locus diseases. *Artif Intell Med*, 2019. **99**: p. 101690.

Annexes

Annexe : Protocole Capture 201 gènes

Protocole de Recherche Biomédicale
N° du protocole

Version CPP n°2.1 du

Identification et contribution de mutations constitutionnelles dans le développement de phénotypes tumoraux extrêmes, de présentation sporadique ou familiale, par une approche de capture ciblée

Titre abrégé : Panel Cancer Genes

Promoteur : CHU de Rouen

Investigateur coordinateur	Equipes participantes
Pr. T. Frebourg	Inserm UMR1079 Unité de Génétique clinique du CHU de Rouen Laboratoire d'Oncogénétique du Centre François Baclesse

Table des Matières

1. Contexte de l'étude	4
2. Objectif de l'étude	6
2.1. Objectif principal	6
2.2. Objectifs secondaires	6
3. Critères d'inclusion	7

1. Contexte de l'étude

Une des plus grandes avancées en cancérologie et en génétique au cours des 20 dernières années fut l'identification **des formes héréditaires de cancer**. Ainsi, trois prédispositions syndromiques majeures peuvent être considérées, de part leur prévalence et leur gravité : (i) la prédisposition au cancer du sein et de l'ovaire, dont les bases moléculaires identifiées sont les gènes *BRCA1*, *BRCA2* et *TP53*, (ii) la prédisposition au cancer colorectal se dichotomisant en syndrome HNPCC lié aux altérations des gènes du système MMR (*MSH2*, *MSH6*, *MLH1* et *PMS2*) et syndromes polyposiques (adénomateux ou hamartomateux) impliquant des altérations des gènes *APC*, *MUTYH* ou *STK11*, *BMPR1A*, *SMAD4* et *PTEN*, enfin (iii) le syndrome de Li-Fraumeni prédisposant à des tumeurs multiples, lié aux altérations de *TP53*. L'identification de ces formes familiales, basée sur le diagnostic moléculaire de ces altérations, participe notablement à l'amélioration de la prise en charge de ces familles, en mettant en place pour les patients et les apparentés asymptomatiques porteurs de la mutation des interventions de dépistage, voire des chirurgies prophylactiques.

Un des arguments les plus évocateurs d'une forme héréditaire de cancer est leur précocité. Or, le bilan d'activité des laboratoires d'oncogénétique en France révèle **qu'une fraction importante de jeunes femmes présentant des cancers du sein précoces (avant 41 ans et même avant 31 ans) n'ont pas de mutations détectables dans les gènes connus faisant l'objet d'une analyse diagnostique, *BRCA1*, *BRCA2* et *TP53*.** De même, un rapport de l'INCa soulignait que parmi les 2000 cas index adressés chaque année en consultation d'oncogénétique pour une suspicion de forme héréditaire de cancer colorectal, le taux de détection de mutations affectant les gènes impliqués connus était inférieur à 15%. Le constat en 2015 demeure donc celui d'une méconnaissance des bases moléculaires des cancers héréditaires.

Compte-tenu de la découverte récente de la mutabilité du génome humain, révélée par les études pan-génomiques, une première hypothèse de travail est **qu'une fraction sous-estimée de cancers sporadiques de révélation précoce soit due à des mutations *de novo*, survenant dans les gamètes parentaux.** Les nouvelles technologies de **séquençage de nouvelle génération (ou NGS)** permettant d'analyser l'intégralité de la région codante des **23000 gènes** humains (ou exome), **une stratégie dite des trios** a été initiée en première approche au sein de l'unité Inserm U1079. Cette stratégie repose sur la soustraction à l'exome d'un cas index (patient atteint d'un cancer précoce sans antécédent familiaux) des exomes parentaux, afin de ne conserver que les variations génétiques survenues *de novo*, que nous supposons donc impliquées dans le phénotype tumoral observé. En 2012, **nous avons pris comme paradigme le cancer épithélial de l'ovaire de forme précoce** puisque ce cancer est surtout fréquent après 50-60 ans et que la survenue d'une forme précoce de l'ovaire est évocatrice d'une mutation constitutionnelle. Chez **une patiente** ayant présenté, à l'âge de **21 ans**, un **adénocarcinome séreux ovarien bilatéral métastatique** en l'absence de toute

histoire familiale de cancer, nous avons identifié à partir des 20000 variations détectées, une **seule mutation *de novo*** (p.Asn386Ser) dans le gène ***INHBA*** codant **la sous-unité β A des inhibines et des activines** qui jouent un **rôle crucial dans le développement ovarien**. Des approches fonctionnelles et des séries de réplication ont permis d'asseoir la contribution de cette mutation *de novo* dans le développement précoce d'adénocarcinomes séreux de l'ovaire¹. Cette même approche a été reproduite chez **2 patientes** ayant développé **un cancer du sein précoce à 26 et 21 ans**, respectivement, conduisant à l'identification dans les 2 cas de **2 mutations *de novo***. Cependant, dans ces deux derniers trios, les mutations *de novo* identifiées et confirmées affectent des gènes peu connus, pour lesquels la démonstration de la causalité par des approches fonctionnelles constitue un écueil. En effet, une des difficultés principales des projets d'exomes même après la filtration bioinformatique des variations identifiées et l'analyse statistique des données est l'interprétation de la conséquence biologique des variations, compte-tenu de la variabilité du génome humain.² Il est en effet difficile pour chaque nouvelle variation identifiée de développer des analyses fonctionnelles *in vitro* ou des modèles animaux pour affirmer ou infirmer l'impact biologique des variations. Or en 2014, il apparaît que **le nombre de gènes dont les mutations somatiques contribuent au cancer est en fait relativement limité**.³ En se basant sur cette observation, nous faisons les hypothèses que : (i) le nombre de gènes dont les mutations constitutionnelles contribuent au déterminisme génétique des cancers est également limité, (ii) les gènes potentiellement altérés à l'état constitutionnel sont identiques à, ou impliqués dans les mêmes voies biologiques que les gènes fréquemment mutés à l'état somatique dans les cancers, (iii) les corrélations génotype/phénotype orientant actuellement les analyses moléculaires peuvent être en défaut (concept d'extension phénotypique), (iv) une fraction de ces altérations moléculaires *de novo* peuvent exister à l'état de mosaïque, dès lors que l'évènement mutationnel survient au stade post-zygotique, et de ce fait uniquement détectable par analyse NGS à très grande profondeur. Dans cette dernière hypothèse, seuls certains tissus de l'organisme sont concernés par la mutation, avec de ce fait une fraction allélique très inférieure à 50% (aux alentours de 5 à 10%) lorsque l'analyse est effectuée à partir d'ADN extrait du sang périphérique. Les nouvelles technologies par NGS à grande profondeur (500 à 1000x) permettent la détection de ces mosaïques, et ont déjà fait l'objet de publications.⁴

Ainsi pour ne se focaliser que sur des variations à forte probabilité de caractère délétère, l'objectif de ce projet est **d'étudier par NGS** non plus l'ensemble des 23000 gènes mais **200 gènes qualifiés de "*cancer genes*"** dont la contribution au cancer a été étayée de façon non ambiguë. L'identification d'une mutation constitutionnelle de l'un de ces gènes sera d'emblé un argument fort de causalité qu'il conviendra d'étoffer par des études de prédiction de l'effet de la mutation et des analyses statistiques. De plus restreindre l'analyse à 200 gènes permettra de disposer d'une **profondeur de lecture très élevée** (chaque base de l'ADN étant séquencée plusieurs centaines à un millier de fois) permettant la détection de mutations *de novo* à l'état de mosaïque.

Nous souhaitons appliquer cette stratégie à **des cas sélectionnés de cancers évocateurs d'un déterminisme génétique du fait de leur précocité (cancer du sein avant 31 ans, cancer de l'ovaire avant 41 ans, cancer colorectal avant 31 ans, cancers pédiatriques)**, dans un premier temps sans antécédents familiaux, **évoqueurs donc de mutations *de novo***, et sans mutation détectable des gènes connus de prédisposition héréditaire au phénotype considéré.

2. Objectif de l'étude

2.1.Objectif principal

- Détermination de la prévalence de variations délétères et/ou probablement délétère à l'**état constitutionnel** parmi un panel de gènes impliqués dans la cancérogénèse, au sein de 5 séries constituées de patients atteints de cancers à développement précoce : cancer du sein avant 31 ans, cancer de l'ovaire avant 41 ans, cancer colorectal avant 31 ans, cancers pédiatriques et cancers rares de l'adulte de développement précoce sans caractère syndromique.

2.2.Objectifs secondaires

- **Interprétation des variations constitutionnelles** délétères ou possiblement délétères et de leur causalité dans le phénotype tumoral observé, par des approches *in silico*, des approches fonctionnelles, des études de ségrégation et des séries de réplication. Si cette interprétation aboutit de manière non ambiguë à une démonstration de causalité, après avis d'expert (RCP d'oncogénétique régionale), le patient sera informé du diagnostic de variant délétère et des conséquences en termes de prise en charge et de conseil génétique.
- Comparer la prévalence de variations constitutionnelles délétères et/ou probablement délétère à l'état constitutionnel parmi un panel de gènes impliqués dans la cancérogénèse, chez les patients présentant un cancer sporadique et chez les patients présentant une histoire familiale de cancer.
- Détermination de la prévalence de variations constitutionnelles délétères ou probablement délétères à l'état constitutionnel, détectables sous forme de mosaïques.

3. Critères d'inclusion

3.1.Série de patients atteints de cancer du sein précoce

- Age supérieur à 18 ans ou accord parental ou des responsables légaux en cas de patient mineur
- Cancer du sein quelque soit le type histologique et le stade, d'**âge de**

survenue inférieur à 31 ans

- Présentation sporadique ou familiale
- Négativité des analyses moléculaires de *BRCA1*, *BRCA2* et *TP53*
- Consentement libre et éclairé

3.2.Série de patients atteints de cancer de l'ovaire précoce

- Age supérieur à 18 ans ou accord parental ou des responsables légaux en cas de patient mineur
- Cancer de l'ovaire quelque soit le type histologique et le stade, d'**âge de survenue inférieur à 41 ans**
- Présentation sporadique ou familiale
- Négativité des analyses moléculaires de *BRCA1*, *BRCA2*
- Consentement libre et éclairé

3.3.Série de patients atteints de cancer colorectal précoce

- Age supérieur à 18 ans ou accord parental ou des responsables légaux en cas de patient mineur
- Cancer colorectal d'**âge de survenue inférieur à 31 ans**
- En cas de suspicion de syndrome HNPCC (présence d'une instabilité micro-satellitaire et/ou extinction d'une protéine MMR en marquage immuno-histo chimique), négativité des analyses des gènes *MSH2*, *MLH1* et *MSH6* par séquençage à haut débit.
- En cas de suspicion de polypose adénomateuse (définie par un nombre d'adénomes>10) ou hamartomateuse, exclusion des formes mendéliennes connues par analyse des gènes *APC*, *MUTYH*, *SMAD4*, *BMPR1A*, *PTEN* et *STK11* par séquençage à haut débit.
- Présentation sporadique ou familiale
- Consentement libre et éclairé

3.4.Série de patients atteints de cancer pédiatrique

- Age supérieur à 18 ans (au moment de l'étude) ou accord parental ou des responsables légaux en cas de patient mineur
- Tumeur solide pédiatrique d'âge de survenue inférieure à 16 ans pour laquelle une recherche de mutation du gène *TP53* a été prescrite (suspicion clinique de syndrome de Li-Fraumeni)

- Exclusion des patients porteurs d'une mutation de *TP53* détectée par les techniques diagnostiques usuelles (Sanger, capture ciblée et QMPSP pour la recherche de réarrangements de grande taille)
- Consentement libre et éclairé

3.5.Série de patients atteints de phénotypes tumoraux extrêmes, sans présentation syndromique

- Age supérieur à 18 ans ou accord parental ou des responsables légaux en cas de patient mineur
- Tumeur solide d'âge précoce et/ou association de primitifs multiples synchrones ou métachrones
- Absence de présentation syndromique

4. Déroulement de l'étude

4.1.Sélection des patients

Depuis 2011, l'Inserm 1079 intègre des chercheurs du laboratoire d'oncogénétique du centre François Baclesse à Caen qui est un des laboratoires français les plus importants et efficaces en terme de diagnostic des formes héréditaires de cancer du sein et de l'ovaire comme l'atteste le bilan d'activité des laboratoires sous l'égide de l'INCa. De plus les consultations d'oncogénétique de Rouen et de Caen ont une forte activité focalisée sur le cancer du sein et de l'ovaire. Nous disposons donc de prélèvements de patientes hautement sélectionnées tant sur le plan clinique que moléculaire, soit en 2015 une série de **81 patientes ayant développé un cancer du sein avant 31 ans sans mutation des gènes *BRCA1* ou *BRCA2***. Pour ces patientes, une première sélection sera effectuée après séquençage (soumis à l'accord des patientes) des gènes ***TP53*, *CHEK2*, *RAD51C*, *RAD50*, *PALB2*, *MRE11A*, *ATM*, *NBS1*, *CDH1* et *BARD1***, gènes inclus dans la capture diagnostique mise au point par le laboratoire d'analyse de Caen mais non analysés en première intention en l'absence d'indications spécifiques portées par l'histoire familiale ou la présentation phénotypique. Ces gènes en effet sont associés à des risques relatifs de survenue de cancer du sein non formellement établis, et ne peuvent donc être utilisés au titre du diagnostic, en dehors de situation clinique spécifique.⁵ Ce premier criblage moléculaire effectué au titre de la recherche est nécessaire pour n'inclure dans le projet *Panel Cancer Gene* que les patientes demeurant sans mutation identifiée après ce criblage étendu.

Notre laboratoire dispose également de données cliniques et de prélèvements ADN de patientes ayant présenté un cancer du sein dont l'extrême précocité avait conduit à une analyse *TP53*, centralisée au laboratoire de génétique

de Rouen. De manière similaire, après accord des patientes, le criblage moléculaire sera étendu afin de ne sélectionner que les patientes sans mutation connue identifiée. A partir de ces deux collections, nous estimons pouvoir sélectionner au moins **40 patientes ayant développé un cancer du sein avant 31 ans de présentation sporadique ou familiale sans mutation détectable** et ayant donné leur accord pour participer à ce projet.

La série de cancers de l'ovaire sera établie de manière similaire, à partir des données cliniques issues des activités des unités de Génétique Clinique de Rouen et de Caen et/ou des familles analysées pour les gènes *BRCA1* et *BRCA2* sur la plateforme de génétique somatique et constitutionnelle de Caen. A partir de cette série, environ 10 cancers de l'ovaire précoces (avant 41 ans) pourraient être inclus.

La série de patients atteints de cancers colorectaux sera sélectionnée à partir des données cliniques issues d'une étude nationale pilotée par l'Inserm U1079 récemment publiée⁶ qui s'intéressait au déterminisme oligogénique de formes évocatrices de cancers colorectaux héréditaires par génotypage de multiples loci polymorphes (SNPs). Pour cette étude ont été recrutés des patients atteints de cancer colorectal ou adénomes en dysplasie sévère, avec critère de précocité ou histoire familiale évocatrice d'une prédisposition génétique. Les syndromes HNPCC et polyposes étaient exclus. Cette collection de patients sera donc filtrée sur le critère d'âge afin de sélectionner les patients atteints avant l'âge de 31 ans.

La série des patients atteints de cancers pédiatriques sera sélectionnée à partir des données cliniques issues de l'activité de l'unité de Génétique Moléculaire du CHU de Rouen, spécialisée depuis plus de 20 ans dans l'identification des bases moléculaires du syndrome de Li-Fraumeni. A ce titre, le laboratoire reçoit l'ensemble des prélèvements adressés pour analyse du gène *TP53* par les oncogénéticiens/oncopédiatres français. Le laboratoire a ainsi contribué à l'actualisation des critères dits « de Chompret », critères nosologiques facilitant la reconnaissance du syndrome de Li-Fraumeni. Le syndrome de Li-Fraumeni présente une forte hétérogénéité phénotypique, avec néanmoins certaines tumeurs extrêmement évocatrices.⁷ Ainsi, certaines tumeurs pédiatriques sont pathognomoniques, notamment les corticosurrénales, les carcinomes des plexus choroïdes et les rhabdomyosarcomes. Néanmoins, chez certains enfants la survenue de ces tumeurs ne s'explique pas par une altération du gène *TP53* et reste à ce jour inexpliquée, justifiant une approche par panel de gènes. Ainsi, nous disposons à ce jour de (i) 20 cas de corticosurrénales précoces, (ii) 30 cas de rhabdomyosarcomes, (iii) 13 cas de carcinomes des plexus choroïdes. Dans un second temps, la série pourra être ouverte à d'autres cancers pédiatriques dont la présentation était évocatrice de syndrome de Li-Fraumeni (ostéosarcomes, environ 60 cas et autres tumeurs cérébrales, environ une centaine de cas).

La série de patients présentant des phénotypes tumoraux extrêmes sans présentation sporadique seront inclus sur proposition des cliniciens généticiens, devant un contexte évocateur d'une altération constitutionnelle causale mais sans élément d'orientation. A ce jour, une dizaine de patients seraient candidats à une exploration par panel de gène, sur des arguments de précocité du diagnostic, ou de multiplicité des primitifs.

4.2. Analyse par NGS

4.2.1. Sélection des gènes

Après revue extensive des données bibliographiques, nous avons établi une liste de plus de 800 gènes impliqués de manière somatique et/ou constitutionnelle dans la survenue du cancer. Ces gènes ont été extraits de plusieurs sources, comprenant (i) les bases COSMIC recensant les altérations somatiques identifiées dans les cancers, (ii) le portail du *Broad Institute* (<http://www.tumorportal.org>) recensant les données issues des travaux de Lawrence et al.³, (iii) les multiples panels académiques ou commerciaux de gènes impliqués dans le développement des cancers ou faisant l'objet d'un ciblage thérapeutique (*gènes drivers*), (iv) le panel diagnostique proposé par l'Inca concernant les formes héréditaires de cancer, (v) les bases OMIM interrogées sur la coexistence de formes syndromiques de cancer et d'altération génétique, ainsi que les publications majeures relatives aux séquençages par NGS des tumeurs (*The Cancer Genome Atlas*). Afin d'obtenir une liste hiérarchisée, nous avons réuni et critiqué les données épidémiologiques, biologiques et fonctionnelles publiées pour l'ensemble de ces gènes, en pondérant les différents arguments. Une liste de 200 *cancer genes* est ainsi retenue pour définir la capture utilisée pour le séquençage.

4.2.2. Aspects technologiques

Pour la capture des 200 gènes ainsi sélectionnés, un design des sondes est obtenu via le logiciel *SureDesign* de chez Agilent. Les coordonnées génomiques de chacun des gènes, et les références de chacun des transcrits de ces gènes décrits sont renseignées. Les exons codants flanqués de 25 paires de bases introniques définissent les séquences d'intérêt, pour lesquelles des sondes sont définies avec un tiling d'au moins 5x. Les régions riches en GC ou connues pour être difficilement capturées sont enrichies en sondes après vérification *in silico* des couvertures obtenues. Une vérification biologique assortie d'éventuelles corrections du design sera effectuée sur les premiers patients pour confirmer la qualité de la couverture (toutes les régions codantes sont bien séquencées) et les profondeurs obtenues (nombre de lecture de chaque base).

Les échantillons ADN des patients inclus seront analysés sur les plateformes de séquençage haut débit de Rouen (Inserm U1079) et de Caen (Laboratoire de Biologie Clinique et Oncologique). Ces plateformes comportent des séquenceurs de nouvelle génération de type NextSeq 500 (illumina). Les librairies d'ADN seront effectuées manuellement ou de manière robotisée selon le protocole SureSelect^{QXT} d'Agilent. Brièvement, les ADN seront fragmentés et indexés par tagmentation (utilisation de transposase) permettant leur séquençage en pool. Les fragments d'intérêt seront capturés par capture en phase liquide à l'aide des sondes Agilent, et séquencées de manière paired-end 2x150pb sur séquenceur NextSeq 500.

4.2.3.Aspects bioinformatiques

De part l'expérience acquise par les deux plateformes en termes de séquençage ciblé et de séquençage d'exomes, des outils bioinformatiques ont été implémentés ou créés pour répondre aux besoins d'analyses bioinformatiques. Deux pipelines bioinformatiques fonctionnent en parallèle (CASAVA et BWA-Picard-GATK) pour effectuer les étapes de *base calling*, *alignement*, *variant calling* et d'annotation des variants. Les variations ponctuelles, les délétions/insertions, les variations du nombre de copies (CNV) et les réarrangements de grandes tailles pourront ainsi être détectés.

4.2.4.Confirmation des variations identifiées

Toutes les variations identifiées au terme du séquençage NGS et de l'analyse bioinformatique feront l'objet d'une analyse en réplicat par une technique validée de type Sanger, SnapShot ou Digital PCR pour les variations ponctuelles, de type QMPST ou CGH array pour les réarrangements de grande taille et les CNV.

4.2.5.Interprétation des variants

4.2.5.1.Interprétation *in silico*

Les variants non-sens (création d'un codon-stop) seront formellement classés dans les mutations délétères (mutation tronquante entraînant le défaut de synthèse de la protéine). Les variants introniques affectant les sites canoniques d'épissage (GT en +1 ;+2 et AG en -1 ;-2) seront également interprétés comme délétères.

Les variants faux-sens (substitution d'un acide aminé par un autre) nécessitent une interprétation biologique, car un certain nombre d'entre elles peuvent être tolérées par la cellule et correspondent bien souvent à des polymorphismes (SNP). Les variations synonymes (modification nucléotidique sans conséquence sur la séquence protéique) sont le plus souvent non délétères. Plusieurs aspects devront être considérés pour cette interprétation :

- *Conservation nucléotidique, conservation acide aminé*

Les variants modifiant une position nucléotidique extrêmement conservée, et/ou un acide aminé hautement conservé plaident en faveur d'une mutation délétère, en considérant que cette conservation lors de l'évolution de l'espèce indique une position critique pour le fonctionnement du gène/ de la protéine. Différents outils informatiques sont disponibles pour apprécier cette conservation

- *Distance physico-chimique*

La substitution d'un acide aminé par un autre peut induire des modifications conformationnelles au niveau de la protéine produite en introduisant des propriétés physico-chimiques différentes au sein de la séquence protéique

(encombrement stérique, hydrophobie/hydrophilie, polarité). Ce paramètre est apprécié par le score de Grantham. Une variation faux-sens conduisant une distance de Grantham élevée plaide en faveur d'une variation délétère.

- *Domaine fonctionnel*

Ce critère étudie la position de la variation au sein de la protéine, en considérant qu'une variation affectant un domaine fonctionnel a une probabilité supérieure d'induire une perte ou un gain de fonction.

- *Scores agrégés*

Ces différents paramètres sont pris en compte dans des logiciels de prédiction *in silico* accessibles en ligne tels que SIFT, A-GVGD, MutationTaster et PolyPhen-2. Ces logiciels sont redondants mais utilisent des algorithmes différents. Lorsque plusieurs d'entre eux concordent en faveur de la pathogénicité, le caractère délétère peut être fortement suspecté.

- *Mutation d'épissage*

Certains variants faux sens exoniques ou introniques autres que les positions canoniques peuvent modifier les sites d'épissage, et par la même la production d'un transcrit. Des logiciels prédictifs de l'impact sur l'épissage sont également disponibles pour appréhender cet effet *in silico*.

4.2.5.2.Interprétation fonctionnelle

L'impact des variations sur le plan fonctionnel sera étudié selon des modèles d'étude adaptés à la nature de la variation (étude de l'ARN messager en cas de variation pouvant affecter l'épissage, étude de la protéine en cas de variation pouvant affecter la traduction ou la fonction de la protéine). Du fait d'une sélection des gènes d'intérêt dans le panel basé sur leurs implications dans le cancer, la disponibilité de modèles d'étude validés (*in vitro* et *in vivo*) et la possibilité de collaboration avec des équipes de recherche focalisées sur ces gènes est attendue.

4.2.5.3.Interprétation épidémiologique

L'existence rapportée de ce variant sera interrogée dans les bases de données recensant les mutations telles que Ensembl, HGMD ou HGVS, ainsi que dans les bases de données locus-spécifiques lorsqu'existantes (telle que LOVD pour *BRCA1/2*).

L'absence ou la faible représentation de ce variant (fréquence de l'allèle mineur inférieur à 0,001) dans les bases de données internationales telles que l'*Exome Variant Server* ou le *projet 1000 genomes*, ainsi que dans les bases de données propres au laboratoire, constituera un argument en faveur de la pathogénicité.

La variation ou une autre altération de ce gène pourra être également

recherchée dans une série de réplication constituée de patients présentant un phénotype proche. Ainsi, dans notre travail séminal concernant l'altération du gène de l'inhibine dans les adénocarcinomes séreux de l'ovaire, l'identification de mutations faux sens dans le gène de l'inhibine ou dans ses partenaires au sein d'une série de réplication constituée de cancers de l'ovaire précoce avait contribué à retenir le caractère délétère de cette altération.

Sous réserve de la disponibilité des familles et de leur accord, des études de ségrégation du variant pourront être effectuées, en recherchant la présence du variant chez les individus atteints dans la famille et en s'assurant de son absence chez les individus sains.

4.2.5.4.Interprétation phénotypique

De part notre sélection initiale des gènes définissant le panel, l'implication de ces derniers dans la survenue d'un cancer est établie, avec toutefois un gradient de conviction. Afin d'étayer la causalité des altérations génomiques, dans le cas d'altération entraînant une perte de fonction, il pourra être recherché dans le tissu tumoral (si celui-ci est disponible) une perte d'hétérozygotie indicative de la contribution du gène au processus de cancérogénèse.

Par ailleurs, l'implication de certains gènes est souvent documentée de manière restreinte à certains cancers. Notre hypothèse relative aux extensions phénotypiques (un gène dont l'altération constitutionnelle est fortement contributive à un type tumoral donné peut également être causal dans un autre phénotype tumoral) nous amènera ainsi à discuter et explorer le mécanisme sous-tendant cette causalité.

4.2.5.5.Cas des mosaïques

Les mosaïques détectables en NGS seront confirmées par une seconde technique de haute sensibilité sur sang périphérique. Dans les cas où serait disponible du matériel histologique, la présence de la mutation sera recherchée au niveau du tissu concerné par le phénotype tumoral (ainsi recherche de la mutation à l'état hétérozygote au niveau de tissu sain mammaire en cas de mosaïque identifiée chez une patiente atteinte de cancer du sein).

4.2.5.6.Synthèse des arguments de causalité

Sur la base de différents arguments précités, chaque variant identifié sera classé selon son degré de pathogénicité. Le modèle de classification des variations génétiques proposé par l'IARC⁸ sera utilisé à cette fin.

4.2.6.Information aux cliniciens et aux patients

Le patient est informé lors de son inclusion du caractère exploratoire de ce projet de recherche. Néanmoins, compte tenu des implications majeures pour sa prise en charge ainsi que celles de ses apparentés, un rendu diagnostique est

prévu dans le cadre de l'étude, dès lors que les conditions sont réunies pour retenir la pathogénicité et l'implication dans le phénotype tumoral de la variation constitutionnelle identifiée dans le cadre de cette étude. Ces conditions seront étudiées au cas par cas par un collège de généticiens, en lien direct avec le généticien clinicien ayant en charge le patient. Le rendu diagnostique au patient sera effectué par le généticien clinicien référent du patient, en respectant les dispositions réglementaires habituelles de confidentialité et en proposant l'accompagnement psychologique nécessaire.

5. Calendrier de l'étude

Le processus de sélection des patients d'intérêt sera achevé en septembre 2015, permettant d'initier les visites d'inclusion pour recueillir les consentements des patients. Les premières analyses seront ainsi débutées lors du troisième trimestre 2015 et se poursuivront sur une période de 12 mois (séquençage groupé de 16 à 32 patients). Les analyses bioinformatiques et l'interprétation des variations identifiées seront effectuées au fur et à mesure.

6. Financement de l'étude

Les analyses étant effectuées sur collections biologiques existantes, aucun coût relatif au prélèvement n'est nécessaire. Les collections étant stockées au niveau des deux plateformes de séquençage (Rouen et Caen), aucun coût de transport n'est également requis.

Les données cliniques relatives aux patients / familles sont connues dès l'inclusion et seront colligées dans un fichier anonymisé. Aucun coût n'est ainsi associé à leur recueil.

Les consommables nécessaires à la préparation des librairies d'ADN et au fonctionnement des séquenceurs représentent l'essentiel des coûts associés à l'étude. Ainsi, les coûts estimés pour 120 patients : (i) réactifs de préparation de librairie 17 500 euros, (ii) réactifs de séquençage Illumina : 20 400 euros, (iii) analyses de biologie moléculaire pour confirmation des variants détectés par analyse ciblée : 5000 euros. L'Inserm UMR1079 dispose déjà à ce jour de 40 000 euros obtenus par cofinancements sur appels à projets auprès du Cancéropole Nord-Ouest et de la Ligue contre le Cancer.