

Approches bio-informatiques de caractérisation d'une signature moléculaire HRD sur des cohortes de patientes atteintes de cancer du sein ou des ovaires en tissus FFPE

Thibault Matis

▶ **To cite this version:**

Thibault Matis. Approches bio-informatiques de caractérisation d'une signature moléculaire HRD sur des cohortes de patientes atteintes de cancer du sein ou des ovaires en tissus FFPE. Sciences du Vivant [q-bio]. 2021. dumas-03411748

HAL Id: dumas-03411748
https://dumas.ccsd.cnrs.fr/dumas-03411748

Submitted on 2 Nov 2021

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

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

université de BORDEAUX

U.F.R. DES SCIENCES MEDICALES

Année 2021 Thèse n° 3226

THESE POUR L'OBTENTION DU

DIPLOME D'ETAT de DOCTEUR EN MEDECINE

D.E.S. de Biologie Médicale

Présentée et soutenue publiquement le 22 octobre 2021

Par **Thibaut MATIS**

Né le 18 février 1992 à Colombes

Approches bio-informatiques de caractérisation d'une signature moléculaire HRD sur des cohortes de patientes atteintes de cancer du sein ou des ovaires en tissus FFPE

Sous la direction du Pr. Nicolas SEVENET

Membres du jury :

Monsieur le Professeur François-Xavier MAHON	Président du jury
Monsieur le Professeur Jean-Philippe MERLIO	Rapporteur de thèse
Monsieur le Professeur Benoît ARVEILLER	Membre du jury
Madame le Docteur Virginie BUBIEN	Membre du jury
Monsieur le Professeur Nicolas SEVENET	Directeur de thèse

Remerciements

Monsieur le Professeur François-Xavier MAHON,

Merci d'avoir accepté de me faire l'honneur de présider ce jury. Veuillez croire à ma profonde reconnaissance.

Monsieur le Professeur Jean-Philippe MERLIO,

Merci de me faire l'honneur de rapporter et de juger mon travail. Vous m'avez ouvert les portes de la biologie moléculaire sur le versant somatique et vous avez fait germer en moi l'envie d'élucider les mécanismes de l'oncogenèse. Merci pour cette formation et les moyens que vous avez mis en œuvre.

Monsieur le Professeur Benoît ARVEILER,

Je suis honoré que vous soyez dans mon jury. Vous m'avez initié à la génétique dans les meilleures conditions et vous avez su répondre à la perfection à mes attentes à la fois professionnelles et pédagogiques. Vous m'avez fait confiance dès la première heure et je vous en suis extrêmement reconnaissant. Je n'oublierai jamais le temps passé au sein de votre service et dans lequel j'ai tant appris.

Madame le Docteur Virginie BUBIEN,

Je suis ravi que tu aies accepté de faire partie de ce jury. Merci pour tout ce que tu m'as appris et pour ta disponibilité. Nos discussions sont toujours sources d'échanges fructueux. Merci de m'avoir accordé ta confiance et d'être à l'écoute de toutes mes attentes.

Monsieur le Professeur Nicolas SEVENET,

Vous avez été présent dès le début de mon aventure dans la génétique et vous m'avez aidé à accomplir des paliers importants dans mes choix professionnels. Vos conseils avisés, votre sagesse, votre notion du détail et de la rigueur, votre écoute et votre humanisme sont autant de qualités que je m'efforce

de suivre au quotidien et que je souhaiterai, un jour, transmettre à mon tour. Vous m'avez fait confiance alors qu'on ne se connaissait pas, puis vous m'avez guidé et poussé à atteindre mes objectifs. Ce travail de thèse en est le résultat. Pour tout cela, je vous en suis extrêmement reconnaissant. Travailler à vos côtés n'est plus un honneur, c'est dorénavant un plaisir. Merci.

Enfin,

Aux Drs. William Foulkes, Paz Polak, Barbara Rivera, et Nancy Hamel, ainsi que toute l'équipe du Will Foulkes Lab avec qui j'ai eu l'honneur de travailler et d'apprendre. Cette aventure internationale bien qu'écourtée à cause du contexte sanitaire, restera pour moi la plus belle expérience humaine qu'il m'a été donné de réaliser jusqu'à présent. La poursuite de nos « *labs meeting* » hebdomadaires est pour moi une ouverture sur le monde qui m'est chère et une opportunité d'enrichissement personnel sans limite. Comptez sur moi pour cultiver ce lien le plus longtemps possible. A José Valenzuela, *my friend* mexicain pour qui le foot et la génétique n'ont plus de secret et qui partage ma passion de l'histoire.

A Julie et Jennifer, bio-informaticiennes de talent qui ont su répondre à mes attentes pour m'aider à monter ce nouveau pipeline de toute pièce et qui ont répondu présent tout le long de ce travail. Vous avez su vous adapter à mon modeste niveau de bio-informatique et fait preuve de patience. Je vous en remercie. J'ai beaucoup appris à vos côtés et j'ai hâte d'en apprendre davantage. Merci Delfine de m'avoir montré les coulisses du NGS et de m'avoir fait profiter de ton expertise sur ce travail. Merci pour ta disponibilité. A Sabrina et Isabelle, merci de m'avoir guidé dans le monde de l'anatomopathologie.

A toute l'équipe de l'unité d'Oncogénétique de l'Institut Bergonié qui m'a accueilli il y a maintenant un an, je tiens à tous vous remercier pour votre gentillesse et votre écoute au quotidien. Travailler avec vous est un véritable plaisir, les choses sont toujours faciles. Vous formez une très belle équipe, et c'est un bel exemple qu'il est possible de travailler ensemble et de se comprendre même lorsque l'on vient tous de différents horizons. A Sabine, Françoise et Natalie qui en plus d'être des collègues de travail formidables, m'ont permis de réaliser ce travail en toute sérénité.

A tous les co-internes avec qui j'ai eu le plaisir de travailler. A Xavier et Cécile L., « mes parents » du premier semestre. Ce stage restera pour moi l'un des meilleurs et vous en êtes en grande partie responsables. A Benoît, tu ne fus malheureusement que de passage mais tu ne réalises pas l'impact que tu as eu sur mes choix et dans ma vie personnelle. Tu resteras un modèle. A Camille et Jacques,

passer ce stage ensemble ne fut que du bonheur, merci pour ces parties de baby-foot et votre soutien. A Vincent, mon unique co-interne de génétique, tu m'as énormément appris. Merci d'avoir toujours été là lorsque j'étais perdu dans ce monde, parfois complexe, de la génétique. A Cécile G., Marine, Thibault, Julie, Constance et Ranou, vous avez été des co-internes formidables que ce soit à l'internat ou dans les labos !

A mes amis rencontrés lors de mon année de master, Adèle, Florian, Lina-Marie et Giorgio, merci pour votre soutien pendant cette année et votre aide tout aussi précieuse.

A toutes les équipes avec qui j'ai eu plaisir à travailler, notamment toute l'équipe de génétique moléculaire de Pellegrin et l'équipe de biochimie de Haut-Lévêque. Quel accueil ! Mention spéciale à Guillaume, Claudio, Julie et Cécile.

A tous les médecins et pharmaciens inspirants avec qui j'ai pu partager des moments précieux, aux Drs. Nicolas Lechevalier, Christine Mouton, Cindy Lauro, Marie-Lise Bats, Laura Zanardo, Laura Blouin, Perrine Pennamen, Aurélien Trimouille, Eulalie Lasseaux, Charline Caumont, Samuel Amintas (merci pour la relecture), David Cappellen et Anika Bensen. Merci pour tous vos conseils, votre confiance, et votre dévouement dans la formation des internes.

Aux amis bordelais, internes ou non (ce n'est pas un défaut), que j'ai eu la chance de rencontrer en arrivant « à la ville » et avec qui je passe des moments toujours aussi magnifiques. Quel chemin parcouru depuis le premier jour ! A Paul, Claire/Caouète, Julie, Gabi, Hadrien, Marin, Antoine, Gabriel, Chloé, Samar, Thomas, Sophie, Mathilde, Alan, et Max. Le temps s'arrête en votre présence.

A la « *triloc* » George Clémenceau. A Olivier/Fratos, merci pour ta présence toujours apaisante et pour m'avoir appris la cuisine. A Théo, merci d'avoir fait raviver la flamme musicale qui sommeillait en moi.

A la coloc rue Huguerie. A Marc, et nos tutos culinaires qui ne vont pas tarder à être enfin reconnus. A Flo, merci de m'avoir fait découvrir Dylan Rocher chaque dimanche de pluie. A Chloé, tu as été présente à une étape importante pour moi, en plus d'être une amie tu m'as ouvert au cercle des généticiens.

A la « *bro-loc* » du Château-Trompette. A Vincent, merci d'avoir visité Sapienza autant de fois avec moi et pour tes conseils souvent très justes. Bientôt nos maisons seront face à face. A Pierre-Louis, le cœur de la mère, merci pour ta gentillesse, ta générosité et ton humour. Tu es un exemple de savoir bien vivre.

A la coloc TMC. A Claire et Marie, merci de m'avoir appris tant de choses simples de la vie qui pourtant n'avaient pas été portés à ma connaissance, et pour ces émissions de qualités. Votre soutien dans cette dernière ligne droite m'a été très précieux. Je vous dois beaucoup.

Aux amis poitevins et ex-poitevins. A Larick, bien qu'étant une poucave et retournant souvent ta petite veste de bobo, merci de me faire profiter de ta sagesse et de ta sérénité depuis la première année. A Xard (ou Xarv), quelle chance d'avoir pu observer un ptérodactyle une fois dans sa vie ! Merci pour ces moments de « retournages » et passés ensemble pendant le confinement. A mes nobles ministres de la Confrérie : J-Chat, Guérem et Gouille, vos titres pompeux ne sont que le reflet d'une amitié solide. Merci de répondre toujours à l'appel pour partir en campagne. A toi Jeune ! (Manu), tu m'as appris à pécher, j'aurais pu m'arrêter là, m'acheter un C15 et me faire appeler Sacoche. Ma vie aurait été tout aussi remplie de bonheur. Enfin à tous ceux avec qui la vie à Poitiers reste l'un de mes meilleurs souvenirs : Bouboule, Clémence, Walking, Coralie, Clément V., Mr Baiz, Ronan, Hélène, Lucas, Antoine, Franco, Stéphane, Julia, Julie, Hugo et JD.

Aux amis des premiers instants, Pish, Jérome, Médérick, Jarod, Kundun, Benji, Rémy, Alex (dit Teuteuil), Anaïs, Maxou (dit Toutouille), Elodie, Guillaume, et Tinmar. Nous n'avons pas pour coutume, nous autres de Saintes, de prendre un peu de temps pour l'émotion et de se dire merci. Je profite donc de ces quelques lignes pour vous partager le bonheur que j'ai de pouvoir compter sur vous depuis bientôt plus de 15 ans et parfois plus. Il me faudrait bien plus qu'une thèse pour dire à chacun l'importance qu'il a eue dans ma vie. Merci.

A Eric D. et Marie-Thérèse H., merci de partager avec moi vos regards sur les différents aspects de la vie.

A tonton Stéphane, tatie Sylvie, Kévin, Nastasia, Françoise, Jean-Claude, Anne, Elodie et Mathilde. A mon oncle Richard, Clara et Nicolas. Vous êtes ma famille, vous êtes ce qui m'est de plus cher. Malgré la distance soyez confiant de l'amour que je vous réserve.

A mes grands-parents. A papi Roger, auquel je pense toujours et qui aurait été fier de voir que je ne pose plus mes coudes sur la table comme à l'usine. A mamie Liliane, de m'avoir ouvert à la culture et m'avoir toujours encouragé dans mes choix. A mamie Blue de m'avoir conté l'incroyable épopée de Thibaut de Champagne avant que je comprenne que nous n'avions rien à voir, lui et moi. Tu m'as appris la simplicité. A papi André, véritable livre d'histoire, tu as su me transmettre ta passion, ainsi que le goût pour les inventions, la curiosité et la construction des cabanes. Tu es responsable en partie de l'homme que je suis devenu.

A Bastien, malgré le fait que tu aies choisi de vivre (pour l'instant) chez nos vieux ennemis, les moments passés avec toi sont toujours d'une grande richesse. Merci de me remettre en question.

A Théophane, bien que je reste le plus grand de nous deux, tu es en revanche un véritable pilier, sur lequel je sais que je pourrais toujours compter. Ce travail en est l'une des nombreuses preuves.

A mon père, tu m'as très tôt initié à la réflexion de comment tourne le monde, et que « si c'est rond, c'est point carré ». Avec toi se tromper n'est jamais une fatalité. Je te dédie ce travail.

A ma mère, tu es celle qui m'apprit l'humilité, l'équité, le partage et la simplicité. Merci de m'avoir forcé à continuer le piano. Je te dédie ce travail.

A Hippolyne, merci de me guider vers la lumière jour après jour.

Table des matières

I. Introduction 14
A. Les formes héréditaires de cancer du sein et l'ovaire 14
1. Généralités 14
2. Modèle des tumeurs héréditaires 14
3. Physiopathologie moléculaire 16
4. Définition des tumeurs HRD 26
5. Mutations constitutionnelles des gènes de la voie HR et leurs conséquences sur l'oncogenèse mammaire et ovarienne 27
B. Intérêt diagnostique des variants des gènes de prédisposition 29
1. Sur le versant constitutionnel, le conseil génétique et la prise en charge chirurgicale 29
2. Sur le versant somatique, la léthalité synthétique des inhibiteurs de PARP 32
C. Diagnostic moléculaire du syndrome HBOC dans les gènes de prédisposition 35
1. La révolution du séquençage par NGS : l'enjeu de la bio-informatique 35
2. Le séquençage en panels de gènes au niveau constitutionnel et/ou tumoral 38
D. Intérêt diagnostique de caractériser l'HRD dans le cancer du sein et l'ovaire 40
1. Le diagnostic moléculaire au-delà des gènes de prédisposition 40
2. Application au versant somatique : essais cliniques 49
3. Application au versant constitutionnel : l'héritabilité manquante des tumeurs HRD 52
E. Stratégie de détection d'une signature HRD par bio-informatique et objectif de thèses 54
II. Méthodes 55
A. Développement d'un pipeline bio-informatique de détection d'une signature moléculaire HRD. 55
1. Arsenal des outils bio-informatiques 55
2. L'analyse en normal/tumeur 68
3. Choix des outils bio-informatiques et mise en place du pipeline 70
B. Sélection des échantillons 74
1. En vue de la validation du pipeline sur une cohorte de patientes atteintes de cancer du sein. 74
2. En vue de l'application du pipeline en routine sur une cohorte de patientes atteintes de cancer de l'ovaire. 75
III. Résultats 77
A. Validation du pipeline 77
1. Démographie mutationnelle constitutionnelle 77
2. Démographie mutationnelle somatique 77
3. Détection des tumeurs HRD et de variants candidats 77

B. Structure du pipeline de routine 78
C. Extrapolation et résultats du pipeline en routine 80
1. Démographie mutationnelle somatique 80
2. Rendement diagnostique de Mutect2 versus STARK 80
3. Détection de la signature HRD 81
4. Détection d'une tumeur *POLE*-mutée 82
IV. Discussion 85
V. Conclusion 91
VI. Références 92
VII. Annexes 102

Table des figures et des tableaux

Figure 1 : Modèle des two hits de Knudson. *15*
Figure 2 : Principales étiologies des 2nd hits. *16*
Figure 3 : Domaines fonctionnels de BRCA1 et BRCA2 (adapté de Roy et al. 2012). *19*
Figure 4 : Principaux acteurs de la voie de la recombinaison homologue et du contrôle du cycle cellulaire (adapté de Nielsen et al. 2016). *23*
Figure 5 : Altérations moléculaires des cancers de l'ovaire séreux de haut grade (d'après Konstantinopoulos et al. 2015). *27*
Figure 6 : Risques cumulés de cancer du sein (à gauche) et de l'ovaire (à droite) chez des patientes porteuses de mutations pathogènes dans les gènes BRCA1 et BRCA2 (d'après Kuchenbaecker et al. 2017). *28*
Figure 7 : Les différentes fonctions biologiques des PARP (d'après Sonnenblick et al. 2014) *33*
Figure 8 : Concept de léthalité synthétique par inhibiteur de PARP dans le cancer de l'ovaire. *35*
Figure 9 : Etapes de la technologie NGS en phase solide (Illumina®) *36*
Figure 10 : Schéma d'un pipeline bio-informatique standard pour l'interprétation de variant et d'altération du nombre de copies par NGS. *37*
Figure 11 : Visualisation des signatures mutationnelles (d'après Koh et al. 2021). *42*
Figure 12 : Profil mutationnel de la signature SBS3. ..... *44*
Figure 13 : Signatures mutationnelles SBS actuellement connues dans les cancers du sein (adapté d'après Nik-Zainal et Morganella 2017). *45*
Figure 14 : Classification des tumeurs HRD à l'aide des signatures mutationnelles (d'après Davies et al. 2017 et Nguyen et al. 2020) *47*
Figure 15 : Représentation schématique des mécanismes des principales cicatrices génomiques de l'HRD (adaptée de Stern et al. 2016) *48*
Figure 16 : Scores de déficience de recombinaison homologue sur différents types d'échantillons de tumeurs (d'après Abkevich et al. 2012) *49*
Figure 17 : Estimation Kaplan-Meier de la survie sans progression chez des patientes atteintes de cancer de l'ovaire séreux de haut grade, en fonction du statut tumoral BRCA et du statut de la déficience en recombinaison homologue (d'après Ray-Cocquard et al. 2019). *51*
Figure 18 : Répartition en pourcentage de l'incidence de l'HRD par type de cancer et des causes de l'HRD chez les patients HRD positif selon CHORD-HRD (d'après Nguyen et al. 2020). *53*
Figure 19 : Etude de l'oncogenèse dans le syndrome HBOC grâce à l'analyse en normal/tumeur en différentes étapes. *54*
Figure 20 : Liste non exhaustive des outils bio-informatiques sous forme de pipeline et selon leur type de détection. *55*
Figure 21 : Etapes de l'oncogenèse (d'après Stratton et al. 2009). *58*
Figure 22 : Chromosome-plot d'un échantillon tumoral généré à l'aide GATK-CNA. *60*
Figure 23 : Principe de détection des CNA et des LOH selon FACETS *62*
Figure 24 : Caractéristiques des mutations somatiques acquises sur des cellules souches humaines (CSH) de différents tissus à l'aide de MutationalPattern. *64*
Figure 25 : Les étapes de SigMA. *67*
Figure 26 : Schéma des différentes analyses du génome tumoral. *68*
Figure 27 : Vignette du pipeline bio-informatique initial développé pour la détection de l'HRD *71*
Figure 28 : Détection des tumeurs HRD en utilisant SigMA et ScarHRD dans la cohorte de cancer du sein. *78*
Figure 29 : Vignette du pipeline bio-informatique développé pour la détection de l'HRD en routine *79*

Figure 30 : Boxplots des scores HRD obtenus à l'aide de SigMA(A) et de ScarHRD(B). 82
Figure 31 : Profil mutationnel de la tumeur T-POS06 correspondant au phénotype mutationnel POLE. 82
Figure 32 : Comparaison du nombre de variants remonté sans et après filtre selon utilisation du mode normal/tumeur ou tumeur-only.. 84
Figure 33 : Scatter-plot entre le nombre de variants (SNV) et le score SigMA-HRD (SigMA_score) issus de 1022 cancers du sein des données du TGCA. 85

Tableau 1 : Avantages et faiblesses de l'utilisation du NGS en pratique courante (adapté de Federici & Silvia Soddu 2020). 40
Tableau 2 : Comparaison des résultats entre l'essai PRIMA et PAOLA-1 évaluant respectivement l'efficacité du niraparib et de l'olaparib en entretien dans le cancer de l'ovaire séreux de haut grade selon le statut tumoral BRCA et le statut de la voie de la recombinaison homologue. 51
Tableau 3 : Liste des différents outils bio-informatiques et fonctionnels de détection de l'HRD, selon le type d'analyse de séquençage 66
Tableau 4 : Comparaison de la sensibilité (Se) et spécificité (Sp) de Mutect2 et STARK selon le type de panel de capture utilisé. 81

Liste des abréviations :

AUC : Aire sous la courbe

ADNdb : ADN double brins

ADNsb : ADN simple brin

BAM : Binary Alignement Map

BER : Réparation par excision de base

CDB : Cassure double brins de l'ADN

CNA : Altération du nombre de copies

CNV : Variation du nombre de copies

CSB : Cassure simple brin de l'ADN

DBS : Substitution à double bases

FFPE : Formol fixé en paraffine

GPC : Gène de prédisposition aux cancers

HBOC : Cancer du sein et de l'ovaire héréditaire

HGSC : Carcinome de l'ovaire séreux de haut grade

HR : Recombinaison homologue

HRD : Déficit en recombinaison homologue

HRP : Recombinaison homologue efficiente

ID : Signature indel

LOH : Perte d'hétérozygotie

LST : Modifications chromosomiques à grande échelle

MMR : Réparation des mésappariements

NER : Réparation par excision de nucléotide

NGS : Séquençage de nouvelle génération

NMF : Factorisation par matrice non négative

PARPi : Inhibiteurs de PARP

PoN : Panel de normaux

RE : Récepteurs hormonaux aux œstrogènes

RGT : Réarrangement de grande taille

RP : Récepteurs hormonaux à la progestérone

RS : Signature de réarrangement

SAM : Sequence Alignement Map

SBS : Substitution de base unique

SNP : Polymorphisme d'un nucléotide

SNV : Variation d'un nucléotide

SSBR : Réparation des cassures simple brin

TAI : Déséquilibre allélique des régions télomériques

VaF : Fréquence allélique d'un/des variant(s)

VCF : Variant Calling File

GPV : Variant constitutionnel pathogène

VUS : Variant de signification incertaine

WES : Séquençage par exome

WGS : Séquençage par génome

I. Introduction

A. Les formes héréditaires de cancer du sein et l'ovaire

1. Généralités

Le cancer du sein est le cancer de la femme le plus fréquent avec environ 58 500 nouveaux cas en France en 2021 (INCa). Plus rare, le cancer de l'ovaire se place à la huitième place avec environ 6 000 cas par an en France. Dans la majorité des cas ces cancers sont sporadiques c'est-à-dire sans composante héréditaire. En revanche, environ 5 à 10% des cas confondus de cancer du sein et de l'ovaire seraient la conséquence d'une prédisposition génétique. La principale étiologie est le syndrome de prédisposition aux cancers du sein et de l'ovaire (syndrome HBOC, pour *Hereditary Brest and Ovarian Cancer*). Ce syndrome est favorisé par des variants constitutionnels pathogènes (GPV) dans les gènes impliqués dans la voie de réparation de l'ADN par recombinaison homologue (HR ou voie HR) : *BRCA1, BRCA2, PALB2, RAD51C,* et *RAD51D*(1). Le syndrome HBOC est une maladie génétique de transmission autosomique dominante mais d'expression récessive, responsable des formes majeures de cancers du sein et de l'ovaire familiales. D'autres syndromes prédisposent aux cancers du sein et/ou de l'ovaire comme le syndrome de Lynch, le syndrome de Li-Fraumeni, les cancers gastriques diffus héréditaires et le syndrome de Peutz-Jegher.

2. Modèle des tumeurs héréditaires

Il est maintenant bien admis que les cancers héréditaires se développent selon le modèle de «two hits» de Knudson(2)(**Figure 1**). Dans les années 70 et basé sur l'expression clinique du rétinoblastome, Alfred Knudson avait supposé que pour qu'une cellule devienne tumorale, celle-ci devait acquérir au moins deux évènements mutationnels initiaux sur le même locus génomique. Ces «*second hit* » peuvent être des événements somatiques, tels que la perte d'hétérozygotie (LOH pour *Loss of heterozygosity*) qui peut être par copie neutre (recombinaison mitotique, disomie uniparentale) ou par perte hémizygote, les 2eme mutations en *trans*, les délétions interstitielles, les conversions géniques, les translocations, ou les mises sous silence épigénétique(3). Le variant hérité est présent dans toutes

les cellules de l'individu (il y a seulement transmission d'un variant, qui, lorsqu'il est délétère, implique une prédisposition tumorale). Dans le syndrome HBOC, la présence d'une copie non-fonctionnelle sur les gènes de la voie de réparation par HR n'est pas suffisante pour altérer son fonctionnement. La voie de réparation par HR est la plus fidèle des voies de réparation des cassures double brins de l'ADN (CDB). L'apparition dans un clone, d'un deuxième évènement provoquant l'inactivation biallélique de ces gènes, engendrera la perte de capacité de réparation des cassures de l'ADN, aboutissant à une accumulation des cassures doubles et simple brins (CSB), et à une instabilité génomique déclenchant le processus de tumorigenèse au sein du tissu mammaire et/ou ovarien(4,5)(**Figure 1**). Il est maintenant bien décrit que le mécanisme majeur d'inactivation biallélique dans les gènes de la voie HR est la perte d'hétérozygotie, par délétion, non-disjonction ou non recombinaison méiotique par exemple(6)(**Figure 2**).

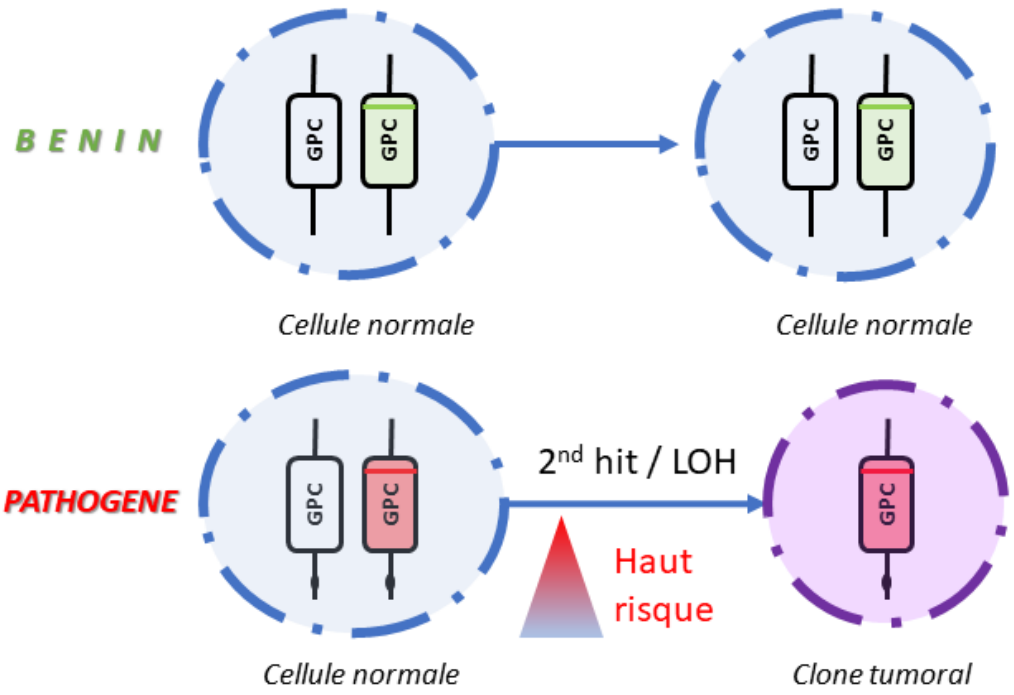


***Figure 1 :* Modèle des two hits de Knudson.** *La présence d'un variant bénin ou d'un polymorphisme dans un gène de prédisposition aux cancers (GPC) n'a pas de conséquence sur l'oncogenèse. En revanche, la présence d'un variant constitutionnel pathogène dans un GPC augmente le risque d'un 2nd évènement de type LOH en trans de ce variant, aboutissant à la perte fonctionnelle de l'allèle normal. Cette haploinsuffisance et la pleine expression du variant pathogène délétère vont favoriser l'émergence d'un clone tumoral.*

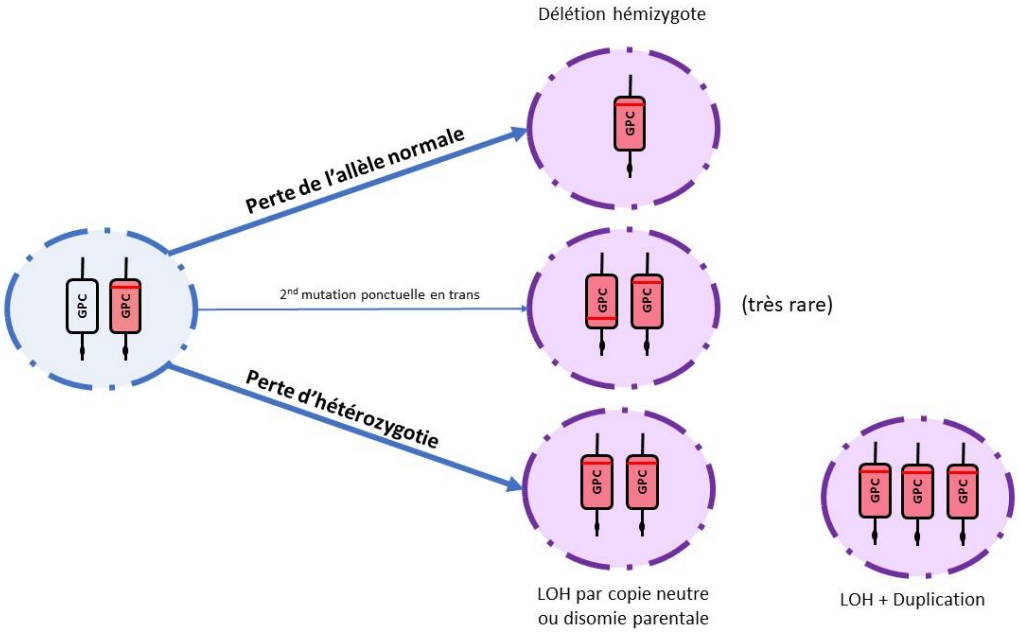


Figure 2 : ***Principales étiologies des 2nd hits.*** *Le mécanisme de 2nd hit le plus commun est la perte en trans de l'allèle normal. Un autre mécanisme moins connu mais fréquent est la perte d'hétérozygotie par copie neutre ou disomie parentale. Ici, l'allèle normal est perdu et l'allèle contenant le GPV est dupliqué. A noter que cet allèle peut être dupliqué n fois. Des mutations pertes de fonctions en trans de l'allèle pathogène sont possibles bien que rares.*

3. Physiopathologie moléculaire

a) Les gènes de la voie de réparation par recombinaison homologue

(1) Les gènes majeurs du syndrome HBOC

Les gènes *BRCA1* (Breast Cancer 1), et *BRCA2* (Breast Cancer 2), sont des gènes suppresseurs de tumeurs de type « caretaker », c'est-à-dire impliqués dans la réparation de l'ADN. Ils ont été identifiés respectivement en 1994(7,8) et 1995(9,10) grâce à des études de clonage positionnel menées dans des familles atteintes de cancers du sein multiples.

Ces gènes sont exprimés de manière ubiquitaire chez l'homme, et jouent un rôle clé dans le maintien de la stabilité génomique, le maintien de la fourche de réplication, la régulation de la transcription, le contrôle du cycle cellulaire et de l'axe mitotique, le remodelage de la chromatine, ainsi qu'un rôle essentiel dans la réparation des CDB de l'ADN par HR(11).

Le gène ***BRCA1*** est localisé sur le bras long du chromosome 17 en 17q21, s'étend sur 80kb et comporte 23 exons. Il code la phosphoprotéine nucléaire BRCA1, composé de de 1863 acides aminés et qui comporte deux domaines fonctionnels majeurs (**Figure 3**) :

- Le domaine RING situé dans la partie amino-terminale, permettant l'interaction avec la protéine BARD1 (BRCA1 Associated Ring Domain 1),
- Le domaine BRCT (BRCA1 C-terminal domain) situé dans la partie carboxy-terminale, permettant l'interaction avec les protéines Abraxas, CtIP (CtBP-interacting Protein) et BRIP1 (BRCA1-interacting protein C-terminal helicase 1).

Le complexe BRCA1-BARD1 permet l'interaction avec RAD51.

Le complexe BRCA1-abraxas est associé au recrutement de BRCA1 aux sites de dommages à l'ADN(12).

Le complexe BRCA1-BRIP1, qui contient également la topoisomérase TOPBP1 (Topoisomerase 2-binding protein 1), est associé à la réparation de l'ADN pendant la réplication(13).

Le complexe BRCA1-CtIP favorise l'activation d'ATR (Ataxia-Telangiectasia and Rad3-related) ainsi que la recombinaison homologue en s'associant au complexe MRN et en facilitant la résection des CDB de l'ADN(14).

BRCA1 possède également un domaine superhélice ou coiled-coil permettant la fixation de la protéine PALB2 (Partner and localizer of BRCA2), facilitant le recrutement de RAD51 aux niveaux des CDB de l'ADN(11), ainsi qu'un domaine SCD (SQ/TQ cluster domain), phosphorylé par ATM. Cette phosphorylation est importante pour l'activation des points de contrôle de la phase G2/M et de la phase S médiée par BRCA1.

Le gène ***BRCA2*** est localisé sur le bras long du chromosome 13 en 13q12-13, et s'étend sur 84kb. Ce gène comporte 27 exons, et code une phosphoprotéine nucléaire, BRCA2, de 3418 acides aminés.

La protéine BRCA2 comporte plusieurs domaines (**Figure 3**) :

- Une extrémité amino-terminale qui interagit avec PALB2,
- 8 motifs BRC dans sa région centrale, permettant de se lier à RAD51,
- Une région DBD (DNA Binding Domain) formée de plusieurs domaines : un domaine H (Helical domain), trois domaines OB (Oligonucleotide Binding) et un domaine T (Tower domain), permettant de se lier à la fois à l'ADN simple brin (ADNsb) et à l'ADN double brins (ADNdb),
- Une extrémité carboxy-terminale contenant un site NLS et un site de phosphorylation par des kinases dépendantes des cyclines (CDK) au niveau de la Sérine 3291 régulant son association à RAD51(11).

Identifiée en 2006 par co-immunoprécipitation de BRCA2(15) , la protéine **PALB2** pour « Partner and localizer of BRCA2 » d'une longueur de 1186 acides aminés, est codée par un gène suppresseur de tumeur *PALB2*, constitué de 13 exons et localisée sur le bras court du chromosome 16.

Cette protéine PALB2 possède au niveau de son extrémité amino-terminale un domaine coiled-coil permettant son interaction avec BRCA1, ainsi qu'un domaine de liaison à l'ADN.

Au niveau de son extrémité carboxy-terminale se situe le domaine WD40, permettant son interaction avec les protéines BRCA2, RAD51, ou RAD51C.

Une fois phosphorylée par ATM, PALB2 présente une fonction dans la stabilisation du filament nucléoprotéique RAD51 et le maintien du point de contrôle G2 après dommages à l'ADN via son interaction avec BRCA2(16).

Les gènes ***RAD51C*** et ***RAD51D*** sont situés sur le bras long du chromosome 17 respectivement en 17q22 et 17q12 identifiés en 1998(17,18). Ils codent des protéines de 376 et 329 acides aminés respectivement qui font partie de la famille des paralogues de RAD51, car elles présentent une homologie de séquence de 20 à 30 % avec RAD51 et entre elles. Les paralogues de RAD51 forment deux complexes distincts : le dimère CX3, composé de RAD51C et XRCC3, et un complexe BCDX2

composé de RAD51B, RAD51C, RAD51D et XRCC2. Bien que leur rôle exact dans le processus de réparation par HR ne soit pas complétement élucidé, ils semblent cependant jouer un rôle important dans la création du filament nucléoprotéique(19). La délétion de l'un des paralogues dans des modèles de souris génétiquement modifiées produit de graves défauts de développement et une létalité embryonnaire précoce avec l'accumulation de dommages de l'ADN(20,21).

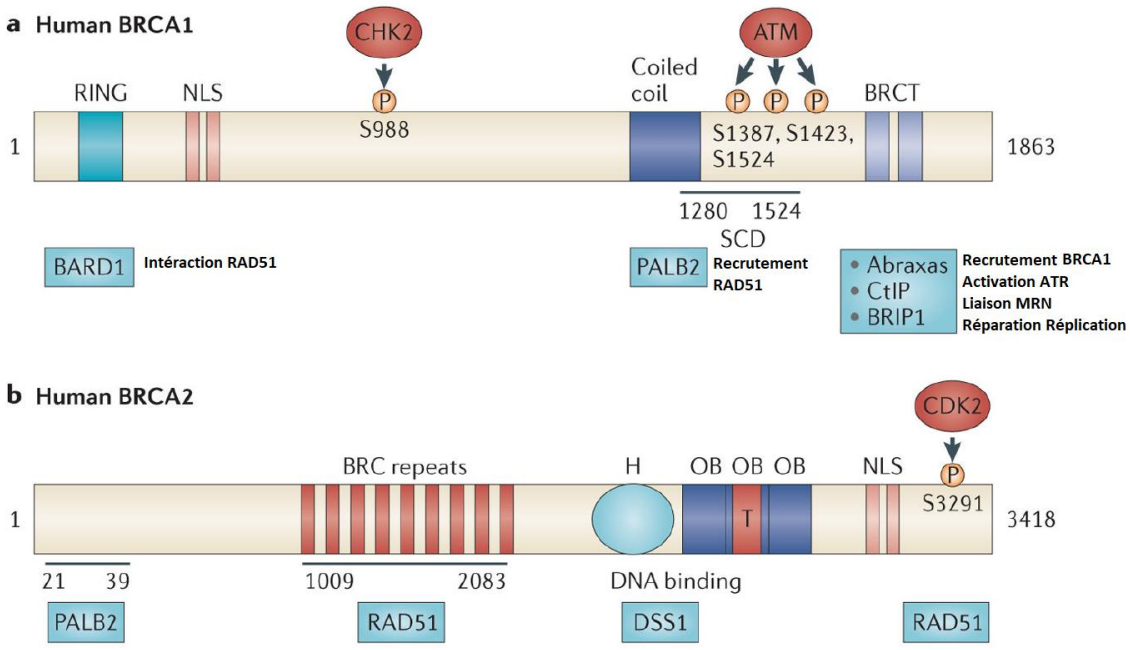


Figure 3 : Domaines fonctionnels de BRCA1 et BRCA2 (adapté de Roy et al. 2012).

(2) Autres gènes impliqués : *ATM, BARD1, BRIP1,* et *CHEK2*

Identifié dès 1995, le gène ***ATM*** pour « *Ataxie-Telangiectasia Mutated* », est un gène suppresseur de tumeur localisé en 11q22.3 sur le bras long du chromosome 11. Il code une vaste protéine de 3056 acides aminées et de 63 exons(22).

La protéine ATM est l'un des senseurs des dommages de l'ADN et joue un rôle critique en phosphorylant et en activant BRCA1 sur le site des cassures(23).

De plus, ATM est impliqué dans la transduction du signal mitogène, la recombinaison méiotique, la détection des dommages à l'ADN et le contrôle du cycle cellulaire.

Identifié à la fin des années 90, le gène ***BARD1*** ou « BRCA1-Associated Ring Domain 1 » est constitué de 13 exons, et est localisé sur le bras long du chromosome 2 en 2q35(24).

Ce gène code une protéine BARD1 de 778 acides aminés qui interagit avec la partie amino-terminale de BRCA1. Associée à BRCA1, BARD1 se lie à l'ADN et interagit avec RAD51. BRCA1-BARD1 renforcent l'activité recombinase de RAD51. D'un point de vue mécanistique, BRCA1-BARD1 favorisent l'assemblage du complexe synaptique et sont indispensables à la stimulation de RAD51(25).

Le gène ***BRIP1*** pour « BRCA1-interacting protein 1 », aussi connu sous *BACH1* « BRCA1 Associated C-terminal Helicase 1 » est un gène constitué de 20 exons localisé sur le bras long du chromosome en 17 en 17q23.2 identifié en 2001(13), et codant une protéine BRIP1 d'une longueur de 1249 acides aminés.

BRIP1 phosphorylé interagit avec le domaine BRCT de BRCA1, afin de contribuer aux fonctions de réparation de BRCA1, mais aussi au contrôle du passage de la phase G2 à M dans un second temps(13).

Le gène ***CHEK2*** pour « Checkpoint Kinase 2 » est un gène constitué de 22 exons localisé sur le bras long du chromosome 22 en 22q12.1, identifié en 1998(26).

Ce gène code une protéine régulatrice du point de contrôle du cycle cellulaire CHK2 de 544 acides aminés. Elle contient un domaine d'interaction essentiel pour l'activation en réponse aux dommages de l'ADN, et est rapidement phosphorylée lors des dommages de l'ADN. Une fois activée , CHK2 est connue pour inhiber la phosphatase CDC25C, empêchant l'entrée en mitose, stabiliser la protéine p53 entraînant l'arrêt du cycle cellulaire en G1 ainsi qu'interagir avec BRCA1 et la phosphoryler(27,28).

b) La voie de réparation par recombinaison homologue

(1) Les cassures double brins de l'ADN

Dans les cellules en prolifération, la voie de de réparation par recombinaison homologue est nécessaire à la réparation des CDB de l'ADN, survenant principalement pendant la réplication de l'ADN(19)**,** et

garant de la stabilité génomique(29). En effet, ces cassures représentent une des formes les plus cytotoxiques de lésion de l'ADN, caractérisées par une cassure simultanée des deux brins d'ADN complémentaires. Ces lésions peuvent être d'origine exogène (toxiques, radiations ionisantes) ou endogène lors d'erreurs de réplication, et sont relativement fréquentes (environ 2 cassures double brins par cellule par heure)(30).

Les protéines BRCA1 et BRCA2 protègent de l'instabilité génétique causée par ces cassures double brins en orientant la réponse biochimique vers la voie HR. Il s'agit d'une voie majeure de réparation de l'ADN très fidèle qui utilise la chromatide sœur homologue comme modèle pour la réparation du brin lésé. La chromatide sœur n'étant disponible qu'après la réplication dans les phases S et G2, la voie HR n'est active que pendant ces phases du cycle cellulaire(19)(**Figure 4**).

(2) La phase pré-synaptique :

- Reconnaissance des cassures double brins :

Lorsque des CDB d'ADN se produisent, les cellules doivent immédiatement réagir afin de les identifier et les réparer. Les extrémités libres de l'ADN sont rapidement détectées par le complexe protéique MRN qui va venir se lier aux extrémités lésées. Le complexe MRN se compose des facteurs MRE11 (également appelé MRE11A), RAD50 et nibrine (NBN ; également appelé NBS1).

Le complexe MRN permet le recrutement de la sérine/thréonine kinase ATM(31). ATM va ensuite phosphoryler et activer plusieurs acteurs de cette voie tels que les protéines du complexe MRN, EXO1, BRCA1, PALB2, ainsi que l'histone H2AX.

L'histone H2AX phosphorylée, communément appelée γ-H2AX, permet le recrutement de BRCA1 et de la recombinase RAD51 afin de former des foyers nucléaires dit « foci RAD51 » sur les sites de lésion.

L'activation d'ATM et la présence du complexe MRN permettent d'initier la résection de l'ADN.

- La résection de l'ADN :

La protéine BRCA1 va ensuite interagir avec deux autres protéines : CtIP et EXO1 (Exonuclease 1), ainsi que le complexe MRN afin de favoriser la résection des extrémités libres de l'ADN(32). Cette résection se fait grâce aux activités nucléases de MRE11 et EXO1.

Le processus de résection finit par exposer d'importantes portions d'ADNsb, lesquelles sont rapidement recouvertes par les protéines du complexe RPA (Replication Protein A), comprenant RPA1, RPA2 et RPA3. Ainsi l'ADNsb, liée aux protéines RPA, ne peut s'apparier avec d'autres ADNsb, empêchant les interactions secondaires parasites issues d'autres processus nucléaires, voire sa dégradation(33).

- L'arrêt du cycle cellulaire :

Le complexe RPA-ADN permet le recrutement du complexe ATR-ATRIP (ATR Interacting Protein). ATR phosphoryle CHK1 qui avec CHK2 inhibent les cyclines permettant l'arrêt du cycle cellulaire en phase S et G2 afin de procéder à la réparation(34).

- Chargement de la recombinase RAD51 :

Le complexe RPA-ADN permet le recrutement de la recombinase RAD51. RAD51 est une recombinase de type RecA qui catalyse l'invasion et l'échange de brins entre des séquences homologues pendant la réparation. Il semblerait ensuite que la protéine BRCA2 à l'aide de la protéine PALB2 interagisse avec les protéines RPA afin de faciliter leur déplacement, permettant de charger la recombinase RAD51 sur l'ADNsb(35,36). La protéine BRCA2 est stabilisée grâce à son interaction avec la protéine PALB2(37). Ce processus est médié par le recrutement des paralogues de RAD51(38).

- La création du filament nucléoprotéique :

Une fois liées , l'ADNsb et RAD51 forment un filament nucléoprotéique nécessaire pour l'appariement avec la chromatide sœur homologue, stabilisé grâce à l'action de la protéine PALB2(39). Cette étape est facilitée par l'action du dimère BRCA1–BARD1.

(3) La phase synaptique :

Il s'agit de l'étape clé de la recombinaison homologue car elle permet au brin d'ADN intact de la chromatide sœur de servir de matrice pour prolonger le brin lésé formant ainsi une jonction de Holliday. Le filament de RAD51 associé à l'ADNsb rejoint l'ADNdb de la chromatide sœur par homologie de séquence, favorisé par la formation des focis RAD51. Lorsque celle-ci est trouvée, RAD51 catalyse l'invasion de l'ADNsb en ouvrant l'ADNdb et formant des boucles de déplacement (D-loop).

(4) La phase post-synaptique :

Le brin néoformé est ensuite prolongé/synthétisé grâce à l'ADN polymérase permettant la jonction avec les extrémités double brins, pour former un produit de réparation en crossing-over ou non(19).

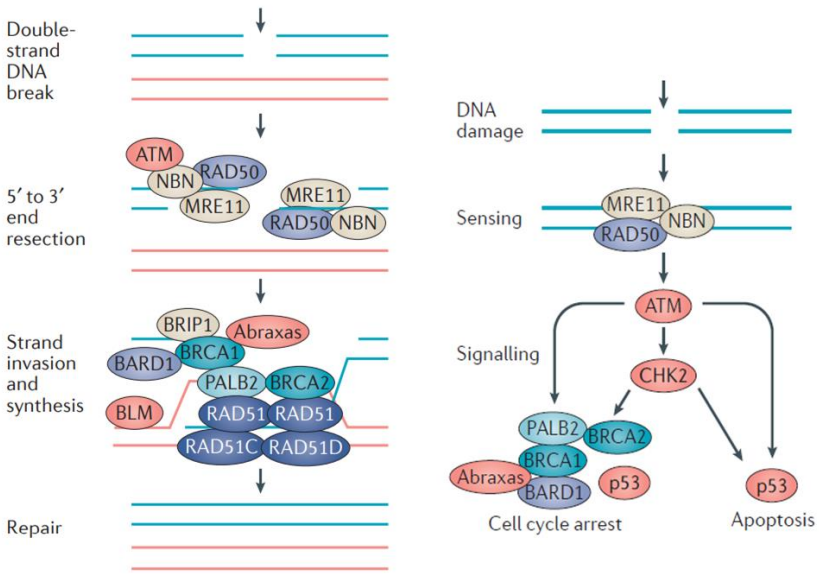


Figure 4 : Principaux acteurs de la voie de la recombinaison homologue et du contrôle du cycle cellulaire (adapté de Nielsen et al. 2016).

Le déficit de la voie de réparation par recombinaison homologue (HRD pour *Homologous Recombination Deficiency*) est un mécanisme augmentant l'instabilité génétique, l'apparition de mutations (charge mutationnelle) et en favorisant ainsi l'activation d'oncogènes et l'inactivation de gènes suppresseurs de tumeurs. Si des mutations pertes de fonctions dans des gènes clés de la voie HR comme *BRCA1*, *BRCA2* ou *PALB2* semblent avoir des conséquences élevées sur la stabilité

génétique(11,40,41), elles semblent plus discutables pour d'autres gènes comme *ATM*, *CHEK2*, ou encore les gènes du complexe MRN ou RAD(6,11,42).

c) Caractéristiques moléculaires des cancers du sein.

Les cancers du sein peuvent être classés en 4 sous-types moléculaires et histologiques distincts(43). Les catégories sont les suivantes :

- Les tumeurs Luminal A représentent entre 60 et 70% des cancers du sein et expriment les récepteurs hormonaux (RE+ et/ou RP+) mais pas Her2 (Her2-). Ces tumeurs sont sensibles à l'hormonothérapie,

- Les tumeurs dites « triples négatives » représentent 10 à 15% des cancers du sein et n'expriment ni les récepteurs hormonaux (RE- et RP-), ni Her2 (Her2-). Ce sous-type est associé au syndrome HBOC et aux mutations constitutionnelles de *BRCA1*(44). Plus de 75% des tumeurs qui apparaissant chez les femmes porteuses d'une mutation *BRCA1* ont un phénotype triple négatif. Ces tumeurs sont malheureusement plus agressives et ne sont ni sensibles à l'hormonothérapie, ni aux inhibiteurs d'Her2 (Herceptin®),

- Les tumeurs Luminal B représentent 10 à 15% des cancers du sein et sont caractérisées par une expression diminuée des récepteurs hormonaux et expriment plus ou moins Her2 (Her2+/-). Ce sous-type est associé au syndrome HBOC et aux mutations constitutionnelles de *BRCA2.* Ces tumeurs restent sensibles à l'hormonothérapie et peuvent être traitées par inhibiteur d'Her2 (Herceptin®),

- Les tumeurs Her2 enrichies représentent 10 à 15% des cancers du sein et n'expriment pas les récepteurs hormonaux mais sur-expriment Her2 (Her2+). Ce sous-type est notamment retrouvé dans les cancers du sein du spectre du syndrome de Li-Fraumeni et aux mutations constitutionnelles de *TP53.* Ces tumeurs sont donc sensibles aux inhibiteurs d'Her2 (Herceptin®).

Les gènes les plus fréquemment mutés et/ou amplifiés dans les cellules tumorales mammaires sont les gènes *TP53* (41% des tumeurs), *PIK3CA* (30%), *MYC* (20%), *PTEN* (16%), *CCND1* (16%), *ERBB2* ou *HER2* (13%), *FGFR1* (11%) et *GATA3* (10%)(45), pour lesquels, hormis *HER2*, il n'existe pas ou peu de thérapies ciblées.

d) Caractéristiques moléculaires des cancers de l'ovaire

Les carcinomes ovariens forment un groupe de maladies hétérogènes avec plusieurs sous-types histologiques, et des caractéristiques clinico-pathologiques et moléculaires distinctes, mais qui peuvent être regroupées en deux types : type 1 et type 2(46). Les carcinomes ovariens de type 1 sont des tumeurs de bas grade, de meilleur pronostic et ne représent que 10% des décès dus aux cancers de l'ovaire. A l'inverse, les carcinomes ovariens de type 2 sont plus agressifs, et sont responsables de 90% des décès dus au cancer de l'ovaire, dont 70 à 80% sont dus à des carcinomes ovariens séreux de haut grade (*HGSC* pour *High Grade Serous Carcinoma*)(47).

Tous sous-types histologiques confondus, les carcinomes ovariens de type 1 présentent plus fréquemment des altérations génétiques dans les gènes *KRAS, BRAF, PTEN, PIK3CA, CTNNB1* et *ARID1A*(48,49).

Il est à noter que les tumeurs à petites cellules de l'ovaire associées à une hypercalcémie (SCCOHT pour *Small-cell carcinoma of the ovary, hypercalcemic type*), sous type particulier dont l'histogénèse demeure inconnue et de mauvais pronostic, sont associées à des mutations constitutionnelles du gène *SMARCA4*(50,51).

A l'inverse des carcinomes ovariens de type 1, les carcinomes ovariens de type 2 présentent presque exclusivement des mutations de *TP53*. Viennent ensuite des mutations moins fréquentes dans les gènes MMR, dans *ARID1A, PIK3CA, KRAS, BRAF, CTNNB1, PPP2CA, FOXM1, AURKA, MTOR, ERBB3, CDK2, BRD4* ou encore *MYC.* A noter que les tumeurs sans mutation *TP53* détectée, présentent des signes de dysfonctionnement de la voie de p53 sous la forme de gain de nombre de copies de *MDM2* ou *MDM4*(48,49).

4. Définition des tumeurs HRD

Les tumeurs sont dites HRD lorsqu'elles présentent une déficience dans la voie de réparation des cassures double brins de l'ADN par recombinaison homologue. La principale cause sont des altérations constitutionnelles ou somatiques dans les gènes de Fanconi ou *BRCA.*

Certaines tumeurs présentent une signature HRD alors qu'il n'est pas retrouvé d'altération dans ces gènes. Ces tumeurs ont des propriétés dites "*BRCAness*", c'est-à-dire des altérations moléculaires qu'elles partagent avec celles qui surviennent chez les porteurs de mutations *BRCA1* ou *BRCA2*. Ces propriétés communes pourraient avoir des implications importantes dans la gestion thérapeutique de ces cancers (voir partie I.D.2)(52).

a) La signature HRD dans les cancers du sein

Environ 20 à 25% des cancers du sein présentent une signature HRD(52–54). Parmi les tumeurs HRD environ 60% sont la conséquence d'inactivation biallélique connue (dans les gènes *BRCA1, BRCA2, PALB2, RAD51C* ou d'hyperméthylation du promoteur de *RAD51C*). Bien que plus rares, 10% des tumeurs HRD peuvent être secondaires à une inactivation biallélique somatique(6,54).

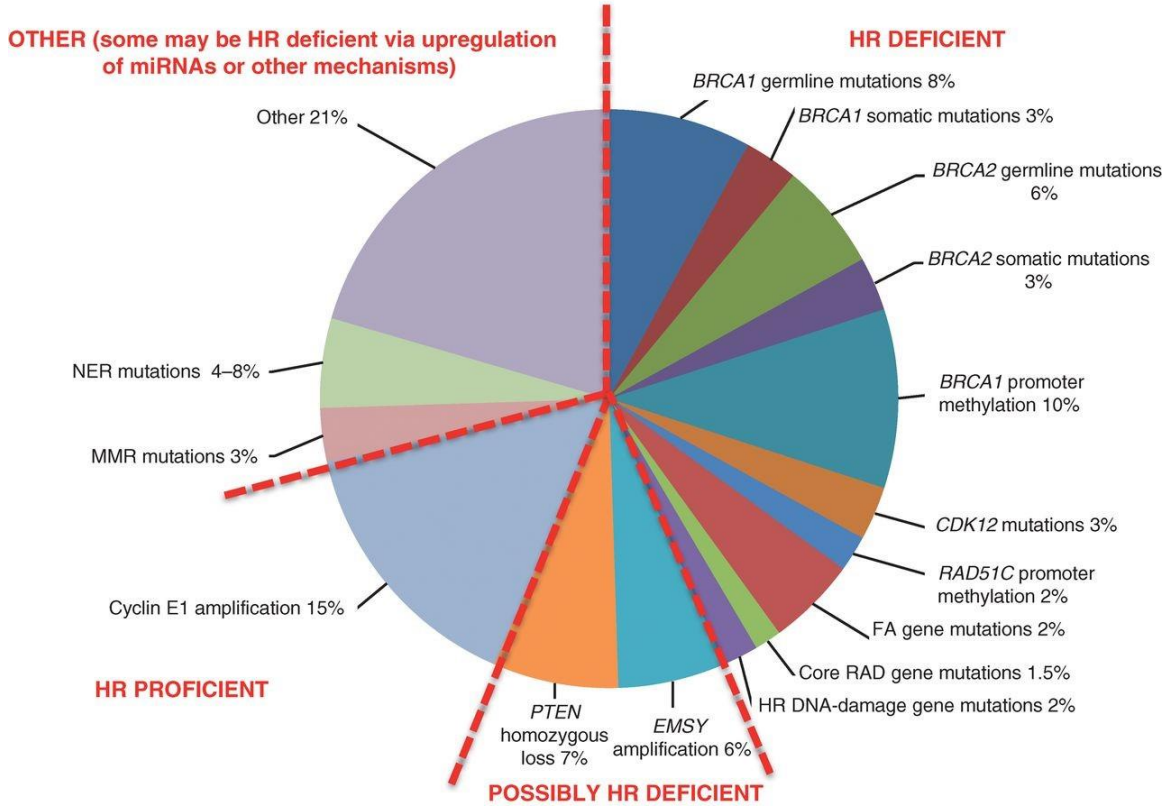
Enfin, il est à noter qu'environ 25% des tumeurs triples négatives, sans atteinte constitutionnelle de *BRCA1,* présentent une hyperméthylation somatique du promoteur de *BRCA1*. Ces tumeurs semblent être de meilleur pronostic(55).

b) La signature HRD dans les cancers de l'ovaire

Environ 50% des HGSC présentent une signature HRD(56)(**Figure 5**). Des mutations constitutionnelles de *BRCA1* et *BRCA2* sont les altérations les plus retrouvées, d'un taux de 14 % à 15 % dans les cancers de l'ovaire en général et jusqu'à 23 % dans les HGSC. Moitié moins fréquentes, les inactivations bialléliques somatiques de *BRCA1* et *BRCA2* sont retrouvées dans 6 % à 7 % des HGSC. Ainsi environ 20% des HGSC-HRD présentent une altération dans les gènes *BRCA1* ou *BRCA2*(56).

L'hyperméthylation du promoteur de *BRCA1* a été identifiée dans environ 10 à 20 % des HGSC-HRD et est mutuellement exclusive des mutations de *BRCA1/2*. A noter qu'une hyperméthylation du promoteur de *RAD51C* survient dans environ 2 % des cas.

Viennent ensuite d'autres altérations de la voie HR comprenant des mutations dans le gène *CDK12*, dans plusieurs gènes de l'anémie de Fanconi (principalement *PALB2, FANCA, FANCI, FANCL* et *FANCC*), dans les gènes RAD (*RAD50, RAD51, RAD51C* et *RAD54L*), et dans les autres gènes impliqués dans la voie HR (*ATM, ATR, CHEK1* et *CHEK2* **(Figure 5)**.



<u>Figure 5 :</u> Altérations moléculaires des cancers de l'ovaire séreux de haut grade (d'après Konstantinopoulos et al. 2015).

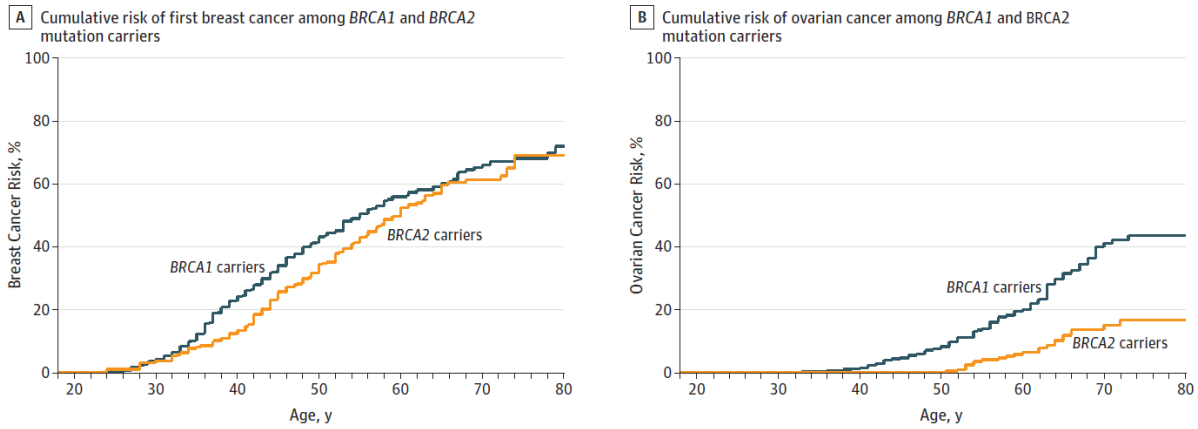
5. Mutations constitutionnelles des gènes de la voie HR et leurs conséquences sur l'oncogenèse mammaire et ovarienne

Les mutations constitutionnelles dans les gènes *BRCA1* et *BRCA2* sont les mutations les plus fréquemment retrouvées et représentent environ 25% des patientes atteintes du syndrome HBOC (57).

Ces gènes sont considérés à haut risque vis-à-vis des cancers du sein et de l'ovaire. En effet, pour les femmes porteuses de mutations pathogènes dans *BRCA1* et *BRCA2*, les risques de développer un cancer du sein à 80 ans sont respectivement de 72 % (IC 95 %, 65 %-79 %) et 69 % (IC 95 %, 61 %-77 %). Moindres, les risques de développer un cancer de l'ovaire à 80 ans sont respectivement de 44 % (IC à 95 %, 36 % - 53 %) et 17 % (IC à 95 %, 11 % - 25 %)(58)(**Figure 6**).

Les patientes présentant des mutations constitutionnelles du gène *BRCA2* sont associées à une meilleure survie globale par rapport aux mutations *BRCA1* chez les patientes traitées par sels de platine. Cette différence serait secondaire à une destruction de cellules cancéreuses plus importante, probablement due à une meilleur sensibilité aux sels de platine(59).

A l'inverse du cancer du sein, la localisation de la mutation est indépendante du risque de développer un cancer de l'ovaire pour *BRCA1* et *BRCA2.* Les mutations en dehors des positions de c.2282 à c.4071 de l'exon 11 de *BRCA1* et les mutations en dehors de l'exon 11 de *BRCA2* (historiquement appelé « OCCR » pour *Ovarian Cancer Cluster Region*) étant associées à un plus fort risque de cancers du sein(58).



<u>Figure 6 :</u> Risques cumulés de cancer du sein (à gauche) et de l'ovaire (à droite) chez des patientes porteuses de mutations pathogènes dans les gènes BRCA1 et BRCA2 (d'après Kuchenbaecker et al. 2017).

Le gène PALB2 est considéré comme étant à risque modéré vis-à-vis des cancers du sein et de l'ovaire. Si le risque absolu de développer un cancer du sein est de 52,8 % (IC 95 %, 43,7 % - 62,7 %) à 80 ans, il est en revanche beaucoup plus modéré pour le cancer de l'ovaire (5% à 80 ans)(60).

Les gènes RAD51C et *RAD51D* sont associés dans une moindre mesure aux syndrome HBOC mais restent des gènes de prédisposition modérée, avec un risque de développer un cancer de l'ovaire à 80 ans respectivement de 11 et 13%(1,60–66).

Bien que faisant partie de la voie HR et semblant apporter un sur-risque dans la prédisposition aux cancers du sein(61,62), les gènes *ATM, BARD1* ou encore *CHEK2* ne semblent pas être impliqués dans la prédisposition aux cancers de l'ovaire(1,63).

B. Intérêt diagnostique des variants des gènes de prédisposition

1. Sur le versant constitutionnel, le conseil génétique et la prise en charge chirurgicale

La découverte d'un GPV impliqué dans les formes familiales de cancer du sein et de l'ovaire dans les gènes de prédisposition présente un intérêt triple :

- Diagnostique : Il permet tout d'abord de poser le diagnostic de cette maladie héréditaire et d'orienter le patient vers une prise en charge et un suivi adapté. Ce premier cas dans la famille est appelé « cas index ».
- Préventif : Il permet ensuite de proposer un test génétique ciblé sur la mutation familiale aux apparentés du cas index afin de pouvoir leur faire bénéficier d'une stratégie de dépistage, voire proposer un traitement chirurgical prophylactique adapté.
- Théranostique : Enfin, depuis récemment, l'identification d'une mutation constitutionnelle dans l'un des gènes *BRCA1/2* offre la possibilité de proposer un traitement par inhibiteur de PARP (voir parties I.B.2. et I.D.2).

Sur le versant diagnostique, plusieurs situations cliniques constituent les indications de recherche de mutations délétères. Elles peuvent être sous la forme d'un score, comme le score d'Eisinger(67) qui comptabilise le nombre de cancer du sein et/ou de l'ovaire dans une famille en prenant compte l'âge d'apparition, et la présence d'une mutation *BRCA1/2* déjà identifiée dans la famille. Mais aussi au

travers d'indications plus spécifique comme c'est le cas à l'Institut Bergonié, où une indication d'oncogénétique est recommandée devant :

- Un cas de cancer du sein avant l'âge de 36 ans ou avant l'âge de 51 ans si triple négatif,
- Un cas de cancer du sein bilatéral dont une atteinte avant 40 ans,
- Un cas de cancer du sein masculin,
- Un cas de cancer épithélial de l'ovaire avant 71 ans ou quel que soit l'âge lorsqu'il s'agit d'un HGSC(68),
- Un cas de cancer du sein et un cas de cancer de l'ovaire chez des apparentés au 1er degré, ou chez la même personne,
- Un cas de cancer du sein et un cas de cancer de la prostate avant 50 ans, ou un cas de cancer du sein et un cas de cancer du pancréas avant 60 ans, liés au 1er ou 2nd degré par l'intermédiaire d'un homme
- Deux cas de cancers du sein chez des apparentés au 1er degré dont l'un avant 40 ans,
- Deux cas de cancers du sein chez des apparentés au 1er degré dont au moins un cas est masculin,
- Trois cas de cancers du sein ou de l'ovaire appartenant à la même branche parentale,

En France et de manière générale, l'analyse est effectuée par séquençage de nouvelle génération (NGS pour *Next Generation Sequencing*) du cas index. Ce séquençage est effectué en panel de gènes (voir partie I.C.2).

Cette analyse permet de rechercher des mutations ponctuelles à l'échelle nucléotidique de type substitution, ou indels, aux conséquences variées : variant non-sens ou faux-sens, décalage du cadre de lecture (*frameshift*), variant d'épissage dans les régions codantes mais aussi aux jonctions exon-intron. Cette analyse par NGS permet aussi la détection d'anomalie de nombre, dit réarrangement de grande taille (RGT).

Eu égard au caractère constitutionnel des variants caractérisés, conformément aux règles de bonne pratique publiées par la HAS et l'ABM, cette recherche doit être toujours confirmée sur un second prélèvement indépendant.

a) Prise en charge des patientes symptomatiques porteuses de mutation constitutionnelle

Les patientes porteuses de GPV dans les gènes *BRCA1/2* et *PALB2,* et ayant développé un cancer du sein, bénéficient d'une surveillance mammaire associant IRM, échographie et mammographie selon deux incidences par sein de façon concomitante, couplées à un examen clinique mammaire, à un rythme annuel, en alternance avec l'examen clinique mammaire annuel (pour une couverture clinique semestrielle). À partir de 65 ans, seule la surveillance mammaire avec mammographie deux incidences/sein (en technique numérique plein champ), échographie et examen clinique, à un rythme annuel et sans limite d'âge suffit.

Il est licite de proposer un geste chirurgical mammaire de réduction de risque (mastectomie bi ou controlatérale) ainsi qu'une ablation chirurgicale des ovaires et de trompes (annexectomie bilatérale) aux patientes traitées pour cancer du sein ou de l'ovaire. Pour le dépistage du cancer des annexes, examen clinique pelvien annuel suffit. En cas d'annexectomie bilatérale : aucune surveillance spécifique n'est recommandée. Compte tenu du pronostic des cancers des annexes, il n'est pas recommandé de réaliser une chirurgie mammaire de réduction de risque chez les patientes ayant eu un cancer des annexes au moins dans les 5 premières années.

b) Prise en charge des patients asymptomatiques porteurs de mutation constitutionnelle

Pour les apparentés asymptomatiques porteuses de GPV dans les gènes *BRCA1/2* et *PALB2*, il est recommandé d'effectuer, vis-à-vis du risque très élevé de cancer du sein, un examen clinique mammaire annuel jusqu'à l'âge de 30 ans (pas de dépistage par imagerie avant sauf si antécédent familial précoce). De 30 à 65 ans ces patientes peuvent bénéficier d'une surveillance par IRM

mammaire plus mammographie et échographie de façon concomitante couplées à un examen clinique mammaire, à un rythme annuel en alternance avec l'examen clinique mammaire annuel seul (pour une couverture clinique semestrielle). Passé 65 ans, seules une mammographie annuelle (une incidence oblique externe) et une échographie suffisent. Pour les hommes, un examen clinique mammaire annuel suffit.

Pour le dépistage du cancer des annexes, examen clinique pelvien annuel suffit. Une annexectomie bilatérale peut être proposée dès 40 ans pour les patientes porteuses de GPV dans *BRCA1,* dès 45 ans pour *BRCA2, RAD51C* et *RAD51D* et 50 ans pour *PALB2* (recommandations Unicancer). En cas d'annexectomie bilatérale : aucune surveillance spécifique n'est recommandée.

2. Sur le versant somatique, la léthalité synthétique des inhibiteurs de PARP

a) La famille des poly(ADP-ribose)

L'accumulation d'erreurs secondaires à l'instabilité génétique provoquée par l'HRD peut être utilisée comme "talon d'Achille"(69). En effet, les thérapies conventionnelles qui ciblent les CDB comme les sels de platine(70–74) ainsi que de nouvelles drogues comme les inhibiteurs de PARP (PARPi)(75–77), sont plus actives lorsque ces protéines réparatrices de l'ADN sont inactives dans une tumeur.

Les poly(ADP-ribose) ou PARP forment une famille de 18 protéines enzymatiques. Leur principale fonction est d'être impliquée dans la réparation des CDB ainsi que des CSB via les voies de réparation par excision de base (BER pour *Base Excision Repair*) et la réparation des cassures simple brin (SSBR pour *Single-Strand Break Repair*). Les membres de cette famille exercent aussi diverses fonctions dans d'autres processus biologiques, comme la régulation transcriptionnelle, la modification de la chromatine, la mitose, l'apoptose, ainsi que le trafic intracellulaire et le métabolisme énergétique (**Figure 7**).

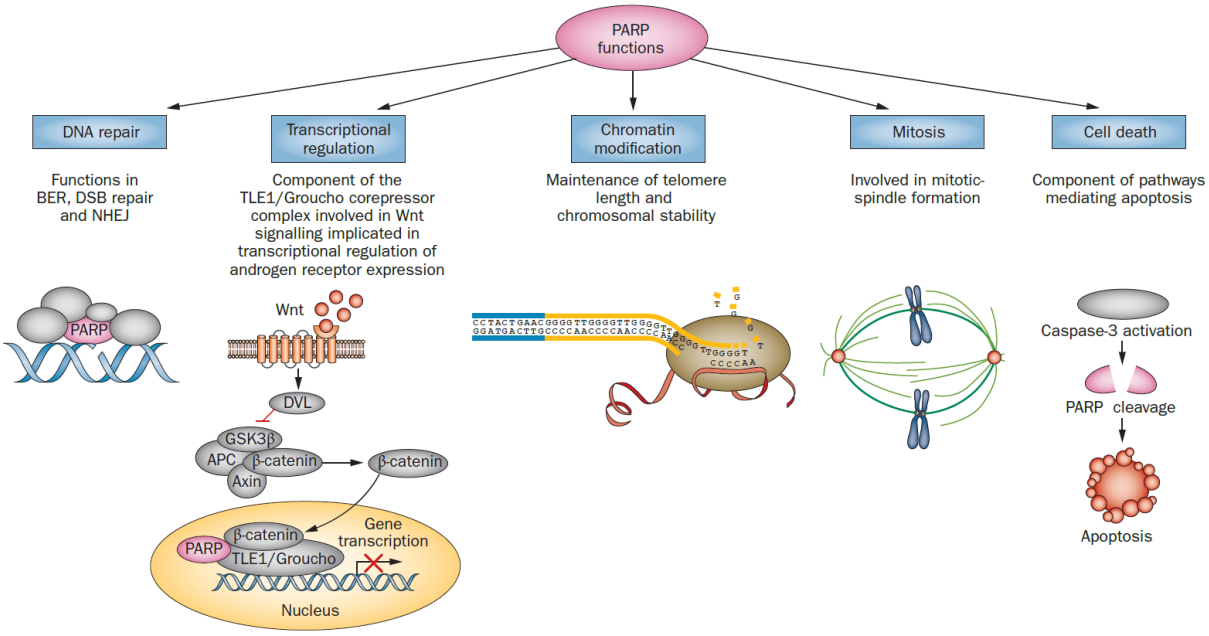


Figure 7 : Les différentes fonctions biologiques des PARP (d'après Sonnenblick et al. 2014)*. PARP1 et, dans une moindre mesure, PARP2 sont importantes dans le maintien de la longueur des télomères et de la stabilité chromosomique. PARP1 fait également partie du complexe corépresseur Groucho/TLE1, et est impliqué comme régulateur transcriptionnel de l'expression du récepteur des androgènes. D'autres PARP fonctionnent dans la réparation des CDB et dans la progression de la mitose (PARP3), et certaines ont des rôles dans la signalisation Wnt et le maintien des télomères (PARP5 et PARP6). PARP1 est également un régulateur de NHEJ, un mécanisme de réparation des CDB. BER : Base Excision Repair ; NHEJ : Non-Homologous End-Joining*

PARP1, la première isoforme découverte, est l'une des protéines les plus abondantes dans le noyau.

D'une taille de 116 kDa, elle est constituée de trois domaines principaux :

- Un domaine amino-terminal contenant le site de liaison à l'ADN via des motifs en doigt de zinc,
- Un domaine central d'automodification ou autoPARylation interagissant avec le domaine BRCT de la protéine BRCA1,
- Un domaine catalytique carboxy-terminal.

Le rôle principal de PARP1 est de catalyser la polymérisation des unités ADP-ribose en une chaîne de 20 à 50 poly(ADP-ribose) dite chaîne PAR, en utilisant le NAD+ comme substrat. Cette chaîne PAR est ensuite fixée sur PARP1 sur son domaine d'automodification, ainsi que sur des histones et des protéines non-histones. Elle permet le recrutement de protéines impliquées dans la réparation de l'ADN en s'y fixant de manière covalente ou non covalente(78).

La fonction catalytique de PARP1 est donc liée à son rôle de senseur de dommages à l'ADN et de molécule de signalisation qui, grâce à son domaine en doigts de zinc, reconnait les CSB et CDB.

b) Rôle dans les cassures simple brin

Les modifications d'une seule base ou d'un nucléotide sont les types de dommages à l'ADN les plus courants ; des milliers de nouvelles lésions sont induites par cellules chaque jour par des métabolites endogènes réactifs mais aussi par la désintégration spontanée de l'ADN. Ces altérations sont réparées par différents mécanismes de réparation : BER, SSBR, mais aussi la réparation par excision de nucléotide (NER pour *Nucleotide Excision Repair*).

c) Le concept de léthalité synthétique de inhibiteurs de PARP

L'inhibition enzymatique de PARP1 dans un contexte d'HRD s'est révélée être une option thérapeutique intéressante. Puisque la protéine PARP1 est essentielle pour la réparation des CSB, empêcher leur réparation les convertiraient en CDB, plus cytotoxiques, et normalement réparées par la voie HR. Ainsi, les cellules ayant une voie HR efficiente (HRP pour *Homologous Recombination Prone*) répareraient leurs CDB, alors que dans les cellules HRD les CSB devenues CDB resteront non réparées et provoqueront une cytotoxicité. C'est ce qu'on appelle le concept de léthalité synthétique (**Figure 8**). Plusieurs molécules dites « inhibiteurs de PARP » ont ainsi vu le jour comme l'olaparib, le niraparib, ou encore le rucaparib, et ont montré une bonne efficacité in vitro sur des lignées cellulaires HRD secondaires à des altérations dans les gènes *BRCA1* ou *BRCA2*(79).

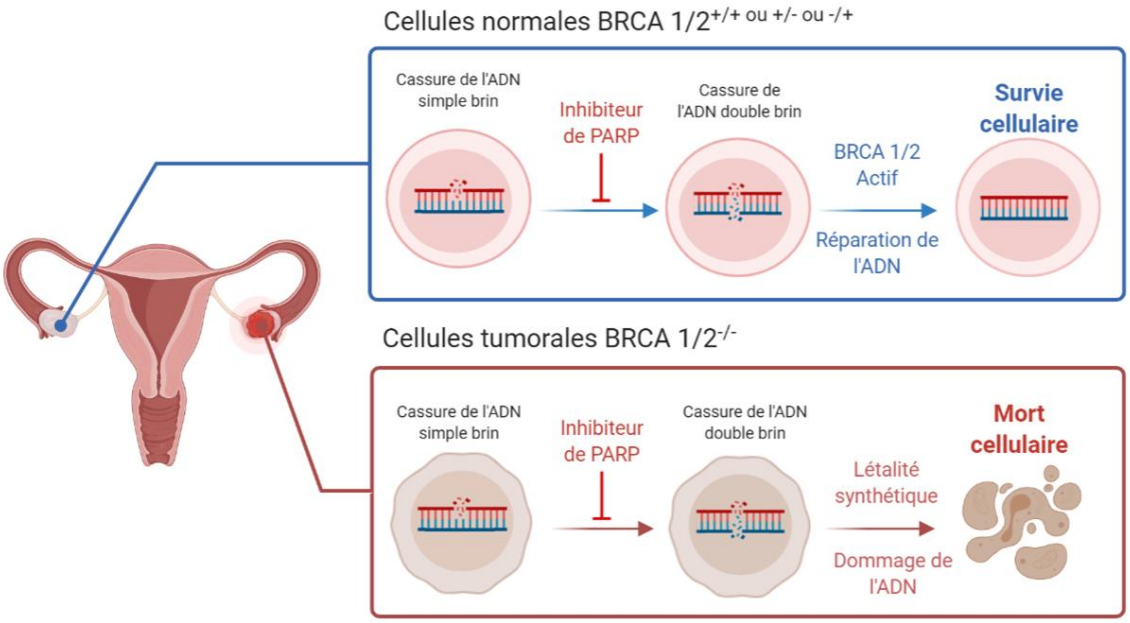


***Figure 8 :* Concept de léthalité synthétique par inhibiteur de PARP dans le cancer de l'ovaire.**

C. Diagnostic moléculaire du syndrome HBOC dans les gènes de prédisposition

1. La révolution du séquençage par NGS : l'enjeu de la bio-informatique

Historiquement, l'analyse moléculaire des premiers gènes de prédispositions décrits (*BRCA1* et *BRCA2*) était généralement réalisée par séquençage Sanger, parfois précédé d'une étape de présélection principalement basée sur la détection d'hétéroduplexes à l'aide de différentes méthodes telles que la chromatographie liquide dénaturante à haute performance (DHPLC), l'analyse de fusion à haute résolution (HRM) ou la technique EMMA. Cette analyse était complétée par le dépistage des RGT, qui est couramment réalisé à l'aide de méthodes semi-quantitatives, telles que la PCR quantitative multiplex de courts fragments fluorescents (QMPSF) ou par amplification multiplex de sondes dépendant d'une ligation (MLPA). Bien que de plus en plus rapide et efficace, l'analyse des gènes *BRCA1/2* était rendue particulièrement laborieuse et coûteuse en temps par le nombre d'amplicons et la nécessité de combiner plusieurs méthodes(80).

Dans les années 2000, l'invention et l'avènement du séquençage de nouvelle génération (NGS), historiquement basé sur des techniques de *polony sequencing*(81), a permis aux laboratoires de

diagnostic moléculaire d'augmenter leur débit, de réduire le délai et d'analyser simultanément différents gènes impliqués dans une maladie spécifique ou un groupe de maladies(80). Les techniques NGS les plus répandues sont basées sur un séquençage parallèle massif après amplification clonale d'ADN en émulsion (IonTorrent®) ou en phase solide (Illumina®) (**Figure 9**).

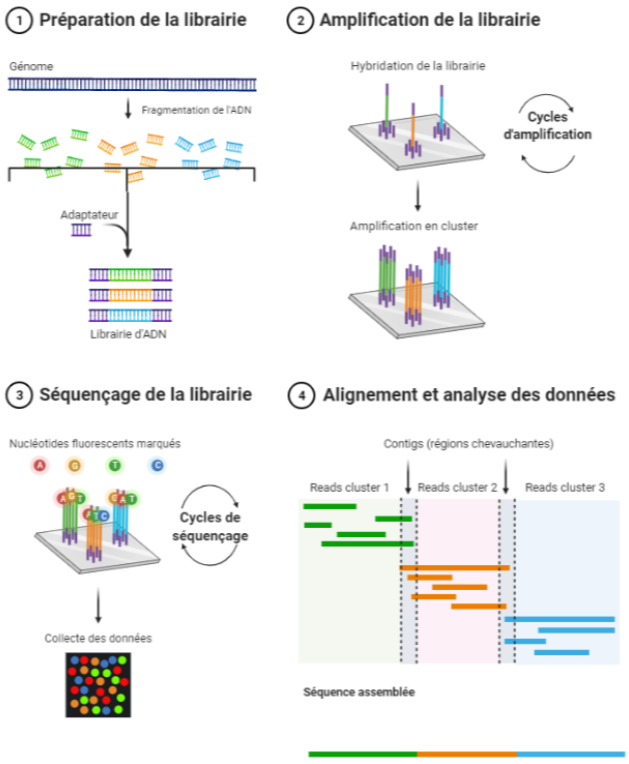


Figure 9 : Etapes de la technologie NGS en phase solide (Illumina®). La préparation des librairies : *les librairies sont créées en utilisant une fragmentation aléatoire de l'ADN puis sont ensuite capturées grâce à des sondes correspondantes aux gènes d'intérêt. Cette étape se termine par la liaison avec des petites séquences spécifiques adaptatrices et contenant un code barre spécifique pour chaque patient.* ***L'amplification :*** *les librairies sont ensuite amplifiées grâce à des méthodes d'amplification clonale par PCR via une hybridation sur une puce contenant des oligonucléotides répartis en cluster.* ***Le séquençage :*** *l'ADN est séquencé par synthèse grâce à l'ajout de nucléotides fluorescents marqués au niveau de chaque cluster. La lumière émise à chaque ajout de nucléotide est recueillie par une caméra.* ***L'alignement :*** *chaque signal est ensuite converti en données informatique. Chaque fragment séquencé est converti en « read ». Les reads sont ensuite alignés sur la séquence de référence*

L'étape supplémentaire par rapport aux techniques expérimentales, est la bio-informatique qui a pour but de traiter, trier et filtrer les données mutationnelles de séquençage afin de les rendre interprétables. La quantité de read générée lors du séquençage, est collectée et stockée sous forme de fichiers FASTQ et nécessite ensuite d'être alignée sur le génome de référence. Une fois alignés (fichiers SAM), ces fichiers sont convertis en fichiers binaires compressés dit fichier BAM contenant

l'information sur l'état de la base (qualité du mapping, variation par rapport à la référence). De ces fichiers BAM, sont ensuite générés des fichiers VCF, au travers d'un « *variant caller* », un programme permettant de ne remonter que les variants d'intérêt. Les fichiers VCF ne contiennent ainsi que les variants que l'on souhaite remonter pour l'interprétation. Cette étape est réalisée en parallèle de « l'annotation » qui ajoute des informations au fichier : type de mutation, conséquence, fréquence allélique, etc. De nombreux autres types de fichiers peuvent être générés à partir de l'un de ces 3 types de fichiers principaux, afin de répondre à diverses exigences (fichiers MAF, fichiers SEG). La bio-informatique s'est donc imposée comme une étape à part entière dans le circuit du diagnostic moléculaire (**Figure 10**).

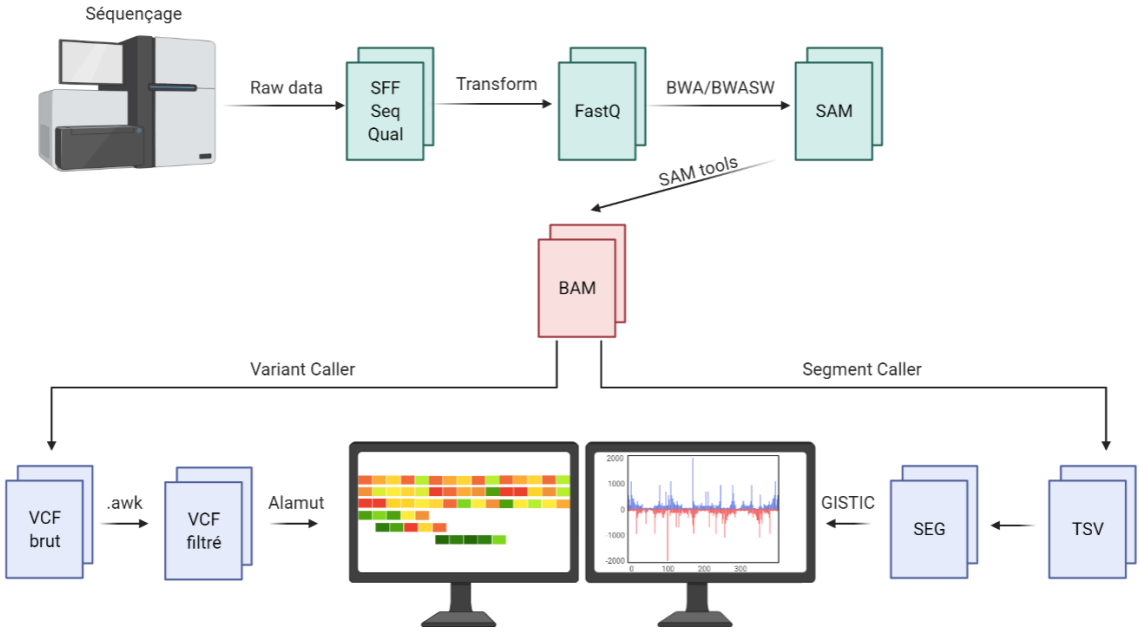


Figure 10 : Schéma d'un pipeline bio-informatique standard pour l'interprétation de variant et d'altération du nombre de copies par NGS. *Les données issues du séquençage sont ensuite transformées en données brutes sous format FASTQ contenant les séquences et diverses informations de qualité. Les fichiers FASTQ sont ensuite alignés et convertis en ficher SAM (Sequence Alignment Map) puis compressés sous format binaire en fichiers BAM (Binary Alignment Map). De ces fichiers BAM peuvent être extraits des fichiers VCF (Variant Calling File), mais aussi des fichiers TSV/SEG contenant des informations quantitatives sur un segment génomique et pouvant être utilisés pour interpréter des altérations du nombre de copies.*

Le principal défi du transfert du NGS vers le diagnostic médical est le développement d'un flux de travail et de pipelines bio-informatiques répondant aux exigences du contrôle de la qualité pour le diagnostic. De nombreux outils bio-informatiques sont désormais disponibles en accès libre sous forme de *Tool Box* (comme GATK), de *Webtools* (comme GenePattern) ou même de pipeline à monter soit

même en ligne (comme Galaxy). En revanche, dans la majorité des cas, la puissance de calcul est à la charge de l'utilisateur, bien qu'il soit possible d'allouer de la puissance sur certaines plateformes (comme la plateforme tout-en-un Firecloud du Broad Institute). Enfin, il existe des prestations complètes, clés en main, commercialisées par diverses sociétés comme Cartagenia ou SeqOne.

Citons par exemple, la *Tool Box* GATK(82), pour GenomeAnalysisToolkit, (Broad Institute), et contenant un ensemble d'outils en en interface par ligne de commande pour l'analyse de données de séquençage à haut débit orientée pour la détection de variants. Ces outils peuvent être utilisés individuellement ou mis à la chaine dans des *workflows* car les résultats obtenus à la fin d'un outil peuvent être directement utilisés par un autre outil de GATK pour des analyses complémentaires et ceci, sans prétraitement ou conversion nécessaire.

Depuis l'émergence du NGS, de nombreux programmes bio-informatiques ont ainsi été développées pour des analyses multiples et variées : analyse d'exome entier (WES pour *Whole Exome Sequencing*), de génome entier (WGS pour *Whole Genome Sequencing*), de RNA-seq, d'altération du nombre de copies (CNA pour *Copy Number Alteration*), de méthylation, de 3D génomes etc…

2. Le séquençage en panels de gènes au niveau constitutionnel et/ou tumoral

En oncogénétique, avant les années 2010 et l'avènement du NGS, seuls les gènes *BRCA1* et *BRCA2* étaient séquencés pour des raisons de temps, de coût et de connaissance étiologique. La capacité de séquençage s'étant grandement améliorée avec le NGS, il est dorénavant possible de rechercher de manière simultanée des altérations dans plusieurs gènes, on parle d'analyse en « panel de gènes »(83).

Les laboratoires ont ainsi pu inclure de nouveaux gènes qui se sont révélés être impliqués dans le syndrome HBOC comme les gènes *PALB2*(84) ou tout récemment *RAD51C/D*(85), ainsi que d'autres gènes syndromiques comme les gènes *TP53*, *PTEN*, *STK11* et *CDH1* qui sont associés à des risques modérés à très élevés de cancer du sein dans les contextes du syndrome de Li-Fraumeni, de la maladie de Cowden, du syndrome de Peutz-Jeughers et du syndrome de cancer gastrique diffus héréditaire.

L'utilisation de panel permet d'inclure d'autres gènes impliqués dans d'autres syndromes de cancers héréditaires d'intérêt diagnostique où des mesures préventives peuvent être mises en place ; c'est le cas des gènes du syndrome de Lynch (*MLH1*, *MSH2*, *MSH6*, *PMS2*), mais aussi d'autres gènes dont l'implication médicale n'est pas encore établie ce jour ou dont la pénétrance est faible dans les familles (*ATM*, *BRIP1*, *CHEK2*, *MRE11*, *RAD50*), mais dont la collecte d'informations permettra de statuer définitivement sur leur implication.

Désormais, la majorité des laboratoires de génétique constitutionnelle mais aussi de biologie des tumeurs utilisent les panels de gènes en routine pour augmenter leur rendement diagnostique dans des maladies héréditaires complexes et hétérogènes, mais également pour la recherche de mutation conductrice éligible à une thérapie ciblée ou d'intérêt pronostique.

Les variants remontés sont ensuite classés en 5 classes selon la probabilité qu'un individu soit porteur d'un changement susceptible de perturber la fonction normale du gène (les variants de classe 5 ayant plus 99% de chances de perturber la fonction normale du gène)(86). Cette interprétation se déroule grâce à l'utilisation de divers critères. Initialement basé sur des données de ségrégation, d'autres critères basés notamment sur des données démographiques, fonctionnelles, ou biochimiques, sont actuellement proposés par l'ACMG(87), le réseau français NGS-Diag ou le réseau anglais ACGS, mais dont l'utilisation reste souvent syndrome dépendant, en prenant en compte la pathogénicité du variant et le phénotype présenté par le patient. En France, et en oncogénétique, seuls les variants de classe 3, 4 et 5 sont rendus dans les résultats transmis aux médecins et aux patients.

Actuellement, l'analyse de 13 gènes (*BRCA1, BRCA2, PALB2, TP53, CDH1, PTEN, RAD51C, RAD51D, MLH1, MSH2, MSH6, PMS2,* et *EPCAM*) est recommandée devant tout contexte évocateur d'une prédisposition héréditaire aux cancers du sein ou de l'ovaire (Unicancer). Cependant, bien que des gènes comme *BRCA1* et *BRCA2* soient bien caractérisés, des variants de signification incertaine (VUS pour *Variants of Uncertain Significance*) continuent d'être identifiés dans ces gènes (environ 7 à 15%)(88), mais aussi dans tous les autres gènes des panels et au fur et mesure que des gènes y sont

inclus (40% dans les panels HBOC(89)). Les VUS sont des variants de classe 3, c'est-à-dire des variants pour lesquels il n'est pas possible à l'heure actuelle de statuer sur leur caractère pathogène (entre 50 et 95% de chances de perturber la fonction normale du gène). Leur fréquence étant logiquement proportionnelle au nombre de gènes inclus dans un panel, reflétant l'un des paradoxes du NGS : l'augmentation de la découverte de nouveaux variants mais dont bon nombre restent sans classification franche (**Tableau 1**).

AVANTAGES	**INCONVÉNIENTS**
Progrès de la technologie à haut débit	Défi pour une utilisation clinique
Utilisation et développement de thérapies ciblées	Identification de mutations sans thérapies ciblées
Caractérisation plus exhaustive des tumeurs	Incertitude quant aux résultats pour les patients
Séquençage simultané de gènes	Coûteuse en temps
Diminution des coûts	Pas toujours accessible
Nombre croissant de données	Interprétation et classification des données difficiles
Nombreux variants découverts	Augmentation du nombre de VUS

***Tableau 1* : Avantages et inconvénients de l'utilisation du NGS en pratique courante (adapté de Federici & Silvia Soddu 2020).**

D. Intérêt diagnostique de caractériser l'HRD dans le cancer du sein et l'ovaire

1. Le diagnostic moléculaire au-delà des gènes de prédisposition

Il existe à ce jour deux grandes catégories de signatures moléculaire de l'HRD : les signatures mutationnelles et les cicatrices génomiques. Les signatures mutationnelles sont le reflet d'une instabilité micro-génomique à l'échelle nucléotidique de l'HRD alors que les cicatrices génomiques sont le reflet d'une instabilité macro-génomique à l'échelle chromosomique. Ces deux signatures sont largement utilisées dans les pipelines bio-informatiques comme outils de détection de l'HRD. Aucune stratégie n'est supérieure à l'autre, ces deux approches ayant démontré la même efficacité (aire sous la courbe (AUC) respectives de 0,837 et 0,823(90)).

a) Les signatures mutationnelles associées à l'HRD à l'échelle nucléotidique

Les signatures mutationnelles sont une analyse du fond génétique tumoral dérivée des changements de bases observés dans une tumeur par rapport à la séquence de référence. Les signatures

mutationnelles sont initialement dérivées des substitutions de base (Single Base Substitution - SBS) en prenant en compte cette substitution dans son contexte tri-nucléotidique, c'est-à-dire les bases jouxtant le variant en 5' et 3' et en prenant comme référence les pyrimidines. On considère qu'il existe 6 types de SBS : C>A, C>G, C>T, T>A, T>C et T>G. Prises dans leur contexte tri-nucléotidique il existe donc 96 variations possibles. Cette analyse en 96 substitutions est particulièrement utile pour distinguer les signatures mutationnelles qui provoquent les mêmes substitutions mais dans des contextes de séquence différents(91). L'application de cette approche sur 30 types de cancer a révélé, en 2013, 21 signatures SBS distinctes dont certaines d'entre elles ont pu être associées à un mécanisme oncogénique connu(92). Par la suite les substitutions à double bases (DBS), les indels (ID) et les signatures de réarrangement (RS) ont été décrites et intégrées dans les signatures mutationnelles (**Figure 11**).

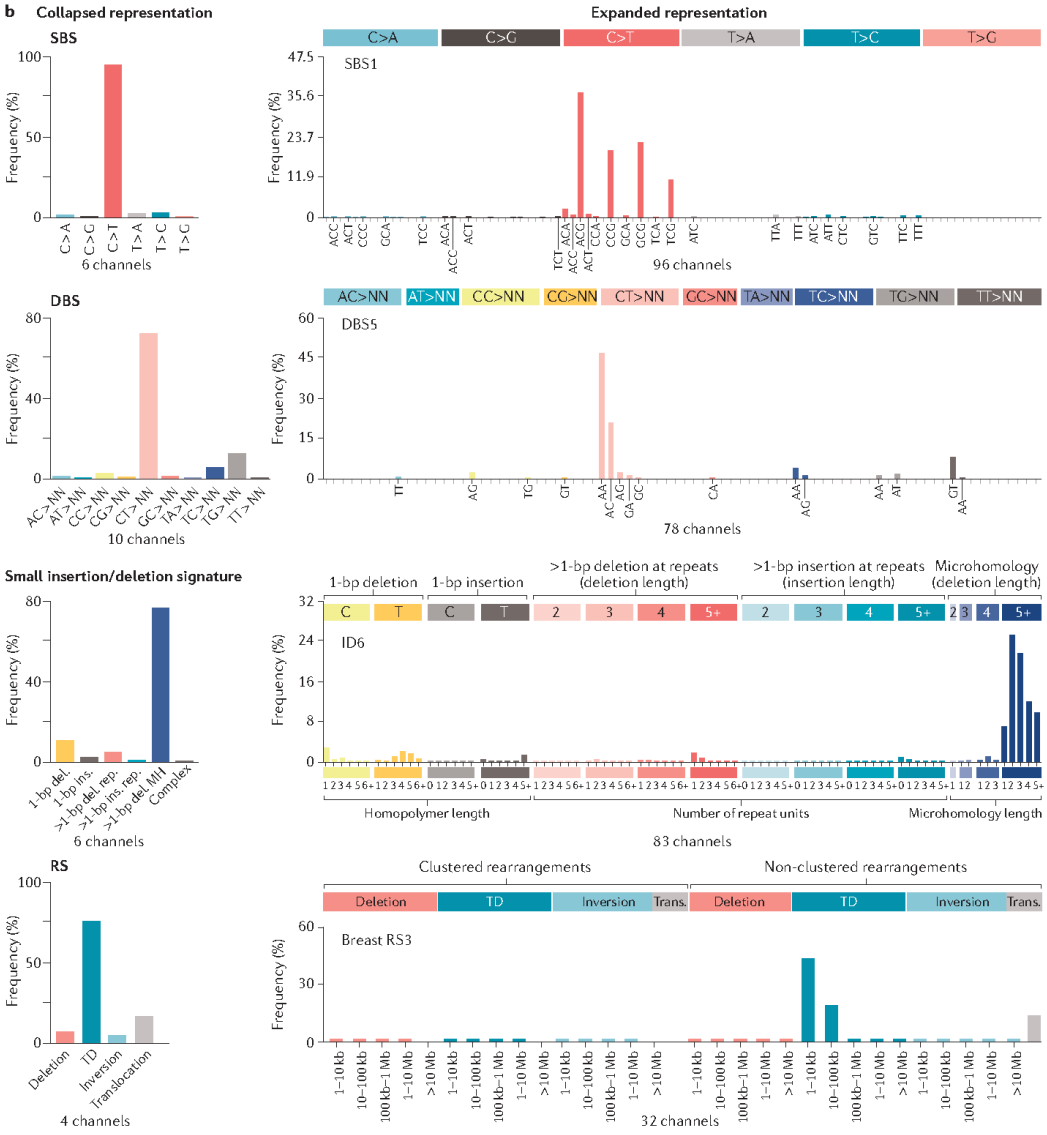


***Figure 11* : Visualisation des signatures mutationnelles (d'après Koh et al. 2021).** *La méthode préférentielle pour représenter les signatures mutationnelles est la représentation en substitution de base unique (SBS) à 96 canaux ou triplets (ici la SBS1). Chaque SBS est ordonnée par classe de mutation (C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, T>G/A>C). L'axe Y rapporte la probabilité qu'une signature génère chacun des 96 triplets. Il est également possible de la réduire à seulement six canaux. Les signatures de substitution à double bases (DBS ; ici la DBS5) peuvent être définies par 78 canaux ou, lorsque la charge est faible, par dix motifs de duplets. Les petites signatures d'insertion ou de délétion (indel de moins de 100 pb), par exemple ici, la signature indel 6 (ID6), sont classées par type (insertion, délétion ou complexe) et - lorsqu'il s'agit d'une seule base comme C ou T - en fonction de la longueur de la chaîne de répétition mono-nucléotidique où elles se produisent. Les indels plus longs sont classées selon qu'elles se produisent sur des répétitions ou qu'elles présentent une micro-homologie aux jonctions d'indel. Les signatures de réarrangement (RS ; par exemple ici la RS3 du cancer sein) peuvent être classées en fonction des quatre types de réarrangements, de la manière dont elles sont regroupées par région et en tenant compte de la taille du fragment réarrangé. del., délétion ; ins., insertion ; MH, micro-homologie ; rep., répétition ; TD, duplication en tandem ; trans., translocation.*

On peut, par exemple, citer les signatures SBS1a et SBS1b qui sont caractérisées par des substitutions C>T dans un contexte NpCpGp, reflétant la désamination spontanée ou enzymatique, liée à l'âge, de la 5-méthylcytosine en thymine et générant des mésappariements G:T. L'absence de détection et de

suppression de ces mésappariements avant la réplication de l'ADN entraîne la fixation de la substitution d'une cytosine par une thymine.

La signature SBS4 est caractérisée par des substitutions C>A dans un contexte NpCpNp. Son profil est similaire au spectre mutationnel observé dans les systèmes expérimentaux exposés aux carcinogènes du tabac tels que le benzo[a]pyrène. La signature SBS4 est donc probablement due à des dommages directs de l'ADN par des mutagènes de la fumée de tabac.

Enfin, la signature 7 est caractérisée par des substitutions C>T dans un contexte TpCpN, reflet des ponts de thymine induits par l'exposition aux UV.

Ainsi la signature SBS1 se retrouve dans bon nombre de cancers liés à l'âge, la signature SBS4 tout particulièrement dans les cancers pulmonaires liés à l'exposition au tabac et la signature SBS7 dans les mélanomes.

En plus des signatures SBS, il est possible de retrouver d'autres sous-types de signatures mutationnelles dans certains types de cancers. Par exemple la signature ID3, caractérisée par des indels d'une cytosine dans des régions d'homopolymère de petite taille est associée à l'exposition au tabac et est retrouvée dans les cancers du poumon. La signature ID13, caractérisée par des indels d'une thymine est associée à l'effets de UV et est retrouvée dans les mélanomes.

En observant les signatures enrichies dans les différents types de cancers, il est apparu qu'une signature était tout particulièrement présente dans les cancers du sein et de l'ovaire liés à des pertes bialléliques de *BRCA1* ou *BRCA2* : la signature SBS3. Cette signature présente un profil avec une représentation plus ou moins égale de l'ensemble des 96 variations, témoin de la réparation défectueuse des lésions de l'ADN en l'absence d'une voie HR efficiente (**Figure 12**). Les signatures SBS8, ID6 (associée à des petites indels de >3bp), RS5 (associée à des délétions plus importantes flanquées de répétitions ou de micro-homologie) et RS3 (associée à la LOH et aux duplications en tandem) sont aussi retrouvées dans les tumeurs associées aux pertes de fonction de *BRCA1/2.* En effet, l'HRD se manifeste principalement par de petites indels et des réarrangements du génome dus à une

réparation anormale des cassures double brins, mais aussi sous la forme de cette signature de substitution de base.

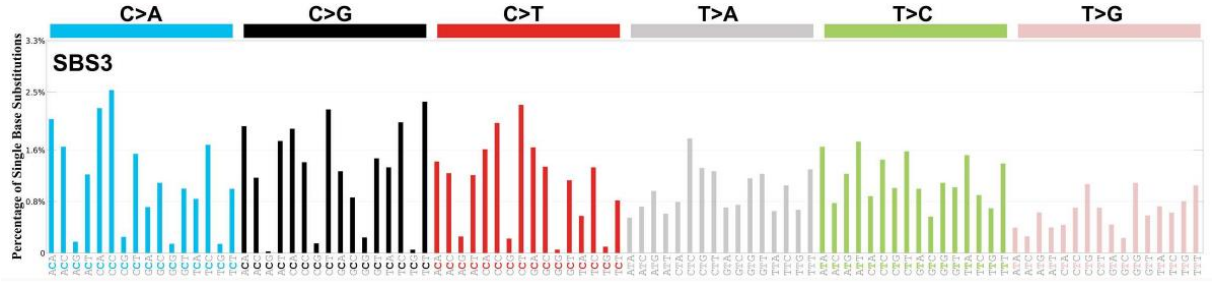
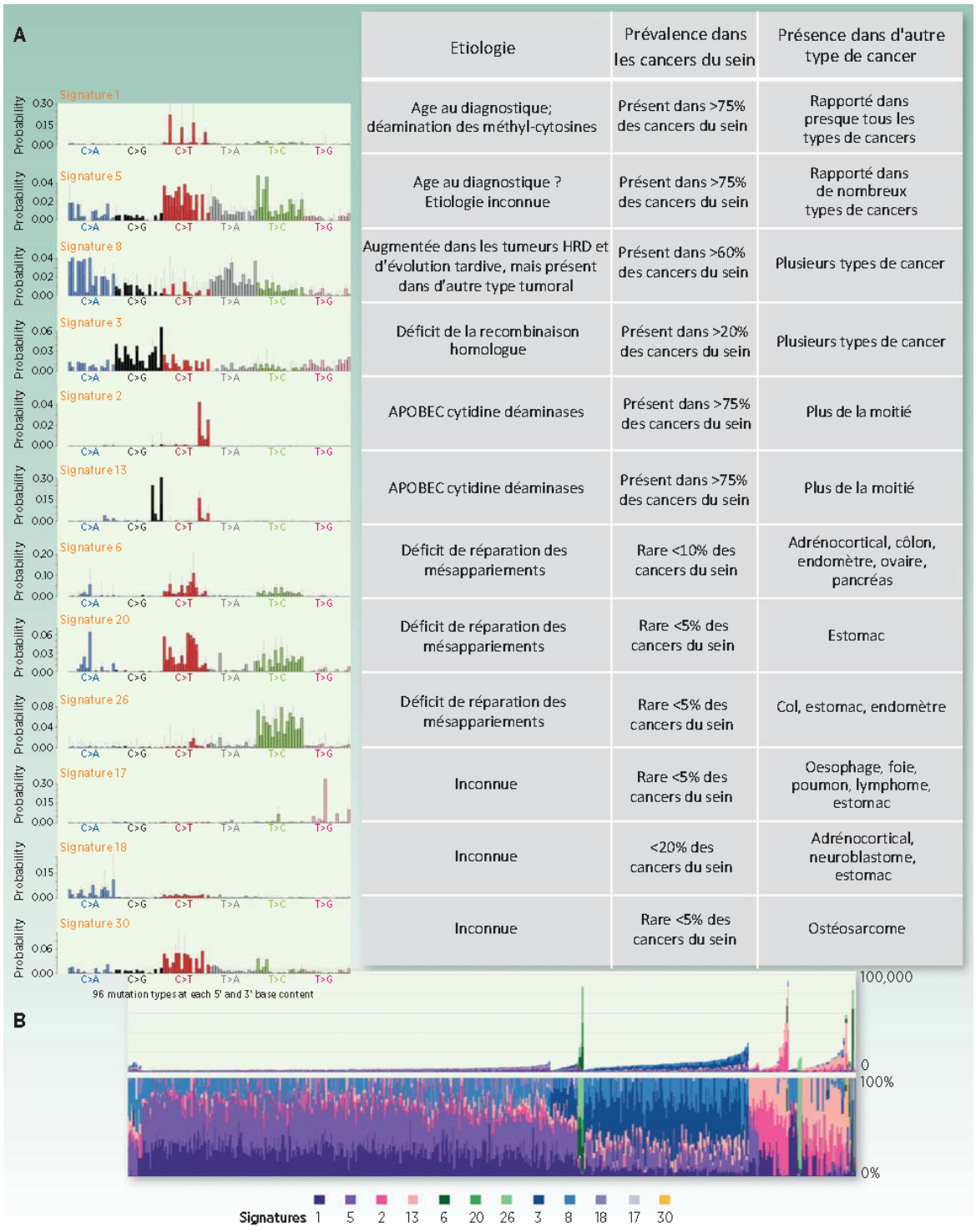


***Figure 12 :* Profil mutationnel de la signature SBS3.** *L'axe des ordonnés rapporte la proportion de chaque substitution dans les 96 triplets (axe des abscisses) que la signature SBS3 génère.*

Alors que certaines tumeurs présentent majoritairement une seule signature reflet d'une cause facilement imputable (la grande majorité des mélanomes ne présente que la signature SBS7, corrélée aux UV(92), d'autres tumeurs peuvent exprimer plusieurs signatures simultanément avec une contribution plus ou moins importante pour chaque signature. On retrouve par exemple, 12 signatures SBS associées aux cancers du sein : les signatures SBS1 et SBS5 qui sont associées à l'âge du diagnostic et qui sont les plus fréquentes ; les signatures SBS2 et SBS13, associées à l'activité des cytidine désaminases APOBEC ; la signature SBS3, associée au déficit en protéines BRCA1/BRCA2 ; la signature SBS8, qui semble être augmentée dans les tumeurs présentant un déficit en protéines BRCA1/BRCA2 (bien qu'elle soit également présente à des niveaux plus faibles dans d'autres tumeurs) ; les signatures SBS6, SBS20 et SBS26 qui sont associées à un déficit de réparation des mésappariements ; et les signatures SBS17, SBS18 et SBS30 qui sont d'étiologie inconnue(45)(**Figure 13**). Il convient de noter que ces signatures mutationnelles ne semblent pas démontrer de spécificité par rapport au sous-type de cancer du sein, qu'il soit classé en fonction du statut des récepteurs aux œstrogènes ou d'autres systèmes tels que le PAM50.



Etiologie	Prévalence dans les cancers du sein	Présence dans d'autre type de cancer
Age au diagnostique; déamination des méthyl-cytosines	Présent dans >75% des cancers du sein	Rapporté dans presque tous les types de cancers
Age au diagnostique ? Etiologie inconnue	Présent dans >75% des cancers du sein	Rapporté dans de nombreux types de cancers
Augmentée dans les tumeurs HRD et d'évolution tardive, mais présent dans d'autre type tumoral	Présent dans >60% des cancers du sein	Plusieurs types de cancer
Déficit de la recombinaison homologue	Présent dans >20% des cancers du sein	Plusieurs types de cancer
APOBEC cytidine déaminases	Présent dans >75% des cancers du sein	Plus de la moitié
APOBEC cytidine déaminases	Présent dans >75% des cancers du sein	Plus de la moitié
Déficit de réparation des mésappariements	Rare <10% des cancers du sein	Adrénocortical, côlon, endomètre, ovaire, pancréas
Déficit de réparation des mésappariements	Rare <5% des cancers du sein	Estomac
Déficit de réparation des mésappariements	Rare <5% des cancers du sein	Col, estomac, endomètre
Inconnue	Rare <5% des cancers du sein	Oesophage, foie, poumon, lymphome, estomac
Inconnue	<20% des cancers du sein	Adrénocortical, neuroblastome, estomac
Inconnue	Rare <5% des cancers du sein	Ostéosarcome

Figure 13 : Signatures mutationnelles SBS actuellement connues dans les cancers du sein (adapté d'après Nik-Zainal et Morganella 2017). A*, Tableau des 12 signatures mutationnelles extraites par factorisation par matrice non négative (NMF). L'extraction des signatures a été effectuée séparément dans 17 types de cancer. Le tableau contient également les étiologies associées à chaque signature, la prévalence de ces signatures dans le cancer du sein, et si la signature est également observée dans d'autres types de tumeurs.* ***B****, Nombre absolu de mutations de chaque signature dans chaque échantillon (en haut) et proportion de chaque signature dans chaque échantillon (en bas).*

Par la suite, il est apparu que la signature SBS3 était significativement associée aux tumeurs liées à des mutations dans des gènes clés de la voie HR comme *BRCA1, BRCA2, PALB2,* ou encore *RAD51C* et que sa caractérisation pouvait être utilisée comme témoin de l'HRD(6,53,54,93). En s'appuyant sur le principe des signatures associés à l'HRD, l'équipe de Serena Nik-Zainal a développé le premier programme de détection de l'HRD basé sur des analyses de WGS et de WES dénommé HRDetect(53). HRDetect identifie les tumeurs déficientes en protéines BRCA1/2 avec une sensibilité de 98,7 % (AUC = 0,98) sur la présence combinée des signatures SBS3, SBS8, ID6, et RS3. L'application de ce modèle à une cohorte de 560 patientes atteintes d'un cancer du sein, dont 22 étaient connues pour être porteuses d'une mutation constitutionnelle de *BRCA1/2*, a permis d'identifier 22 tumeurs supplémentaires avec une perte somatique de BRCA1/2 et 47 tumeurs avec une déficience fonctionnelle de BRCA1/2 pour lesquelles aucune mutation n'a été détectée (**Figure 14**). Plus récemment, une équipe du Hartwig Medical Foundation a développé CHORD, un programme permettant de classifier les tumeur HRD par *machine learning*(54). CHORD a été initialement entraîné à détecter les différentes conséquences de l'HRD à l'échelle nucléotidique que sont les SNV, les indels et les réarrangements de grande taille sur des tumeurs présentant une mutation délétère de *BRCA1* et *BRCA2,* puis à générer un score HRD. De plus, CHORD a été entraîné à classer les tumeurs selon que les profils mutationnels étaient dus à des pertes de fonction dans *BRCA1* (BRCA1-type HRD) ou dans *BRCA2* (BRCA2-type HRD). Grâce à cette classification, il semble que les tumeurs avec des pertes de fonction dans *PALB2* ou *RAD51C*, partagent le même profil que les tumeurs BRCA2-type HRD (**Figure 14**).

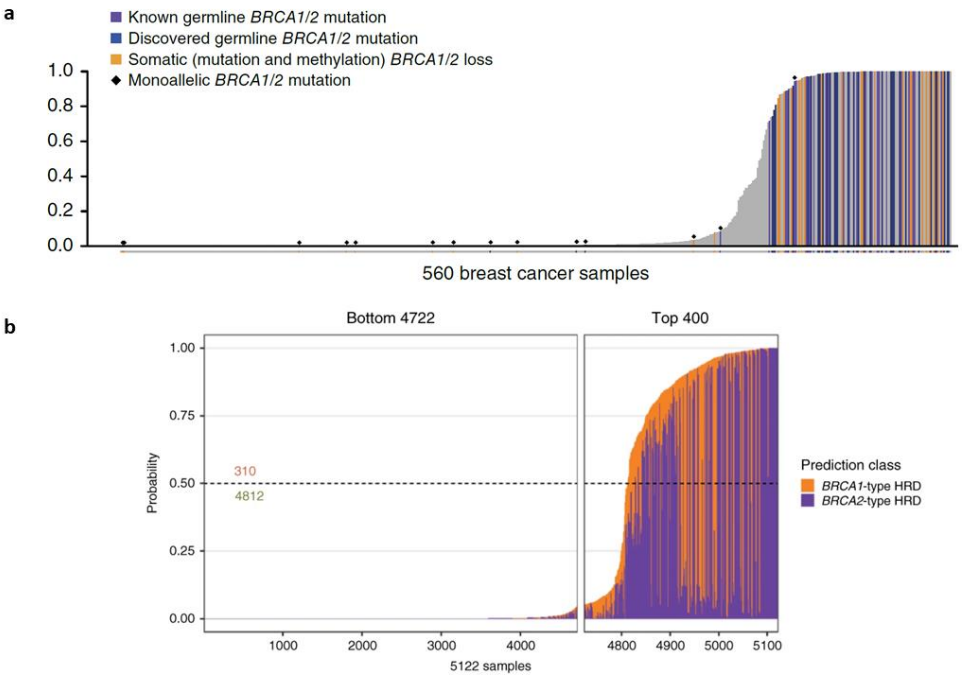
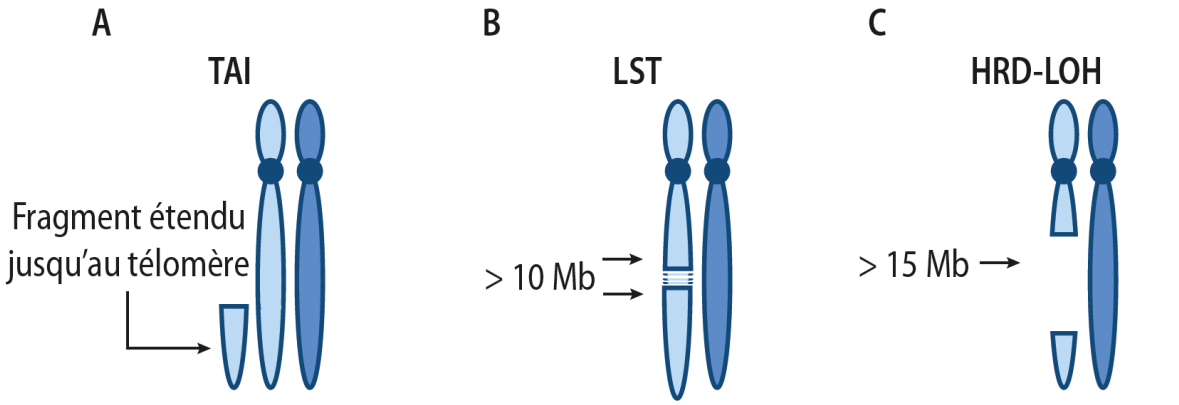


Figure 14 : Classification des tumeurs HRD à l'aide des signatures mutationnelles (d'après Davies et al. 2017 et Nguyen et al. 2020)*. **a**. Score HRDetect de 560 échantillons de cancer du sein, classés de gauche à droite du plus faible au plus élevé. Les barres colorées représentent à la fois les échantillons présentant des mutations monoalléliques et ceux présentant une perte du second allèle : violet, porteurs précédemment connus de mutations constitutionnelles BRCA1/BRCA2 ; bleu, porteurs nouvellement découverts de mutation constitutionnelles BRCA1/BRCA2) ; orange, porteurs de mutations constitutionnelles somatiques de BRCA1/BRCA2 et d'une hyperméthylation de l'ADN du promoteur de BRCA1) ; les losanges noirs au-dessus des barres indiquent les échantillons avec des allèles mutants constitutionnels monoalléliques de BRCA1/BRCA2 qui ont conservé le second allèle.**b**. Score HRD par CHORD, le diagramme à barres montre la probabilité HRD (score > 0,5) pour chaque patient (hauteur totale de la barre), chaque barre étant divisée en segments indiquant la probabilité HRD de type BRCA1 (orange) et de type BRCA2 (violet). 310 patients ont été prédits HRD tandis que 4812 ont été prédits HRP.*

b) Les cicatrices génomiques de l'HRD à l'échelle chromosomique

Plusieurs conséquences de l'HRD à l'échelle chromosomique ont été mises en évidence en 2013 grâce à des études de SNP-array : le déséquilibre allélique des régions télomériques (TAI pour *Telomeric Allelic Imbalance*), les modifications chromosomiques à grande échelle (LST ou *Large Scale Transition*) et la perte d'hétérozygotie (LOH pour *Loss of Heterozygosity*)(**Figure 15**).



TAI = Telomeric Allelic Imbalance ; LST = Large-Scale Transitions ; HRD-LOH = Homologous-Repair Deficiency

Figure 15 : ***Représentation schématique des mécanismes des principales cicatrices génomiques de l'HRD (adaptée de Stern et al. 2016).***

(1) Le déséquilibre allélique des régions télomériques ou TAI

Les défauts connus de la réparation des cassures double brins de l'ADN, y compris la perte de la protéine BRCA1, provoquent la formation spontanée de structures chromosomiques triradiales et quadriradiales, qui sont des indications cytologiques de recombinaison chromosomique aberrante. La résolution de ces réarrangements chromosomiques lors de la mitose peut donner lieu à de vastes régions de déséquilibre allélique et/ou de modifications du nombre de copies s'étendant au télomère (TAI). Sachant qu'un défaut de la voie HR causé par le dysfonctionnement de *BRCA1* ou *BRCA2* entraîne une sensibilité aux sels de platine, Birkbak et al. ont analysé sur des lignées cellulaires si la présence de TAI survenant dans un contexte d'HRD pourrait également être associée à la sensibilité aux sels de platine dans les cancers(94). Les TAI étaient associés à la réponse au traitement par sels de platine suggérant que la charge de cette anomalie génomique était le témoin d'une déficience sous-jacente de la réparation de l'ADN dans le sous-ensemble sensible aux sels de platine de ces cancers.

(2) Les modifications chromosomiques à grande échelle ou LST

En se basant sur le fait qu'un déficit de la voie HR induisait des cassures chromosomiques, Popova et al. ont démontré que des ruptures chromosomiques entre des régions adjacentes d'au moins 10 Mb, étaient un indicateur robuste du statut mutationnel *BRCA*(95). La majorité de ces LST sont des translocations inter-chromosomiques et leur distribution est enrichie dans des régions riches en gènes.

Dans des lignées cellulaires de cancer du sein *BRCA* muté, le taux de LST était corrélé à l'absence de formation de foci RAD51, et donc d'une absence de réponse de la voie HR.

(3) La perte d'hétérozygotie ou LOH

Sur la même hypothèse de l'instabilité génétique causée par l'HRD, Abkevich et al. ont analysé la distribution et la taille des LOH dans des cohortes de cancers de l'ovaire porteurs d'altérations dans les gènes *BRCA1/2* et *RAD51C*(96). Le nombre de régions avec des LOH de grande taille (>15Mb mais inférieure à un chromosome entier) était significativement plus élevé dans les tumeurs *BRCA1* ou *BRCA2* déficient (**Figure 16**). A l'inverse, le nombre de régions avec des LOH emportant l'ensemble d'un chromosome était significativement plus élevé dans les tumeurs sans altération *BRCA1* ou *BRCA2*. À partir de ces échantillons, les auteurs ont déterminé un score de déficience de recombinaison homologue défini comme le nombre de régions LOH >15 Mb, au sein d'un génome tumoral. Les tumeurs HRD avaient un score HRD plus élevé que les tumeurs HRP.

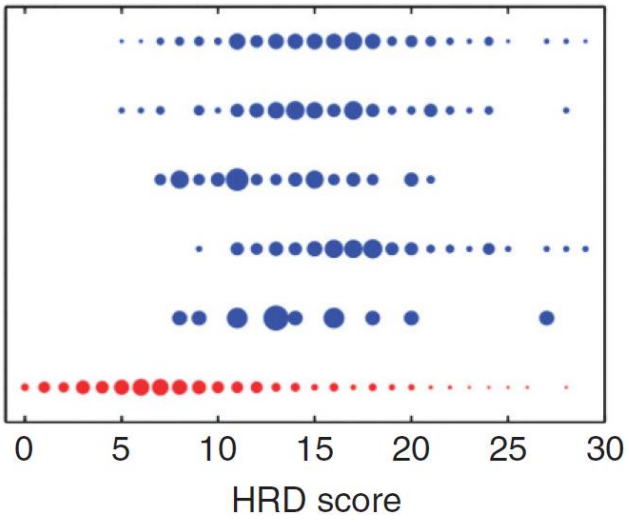


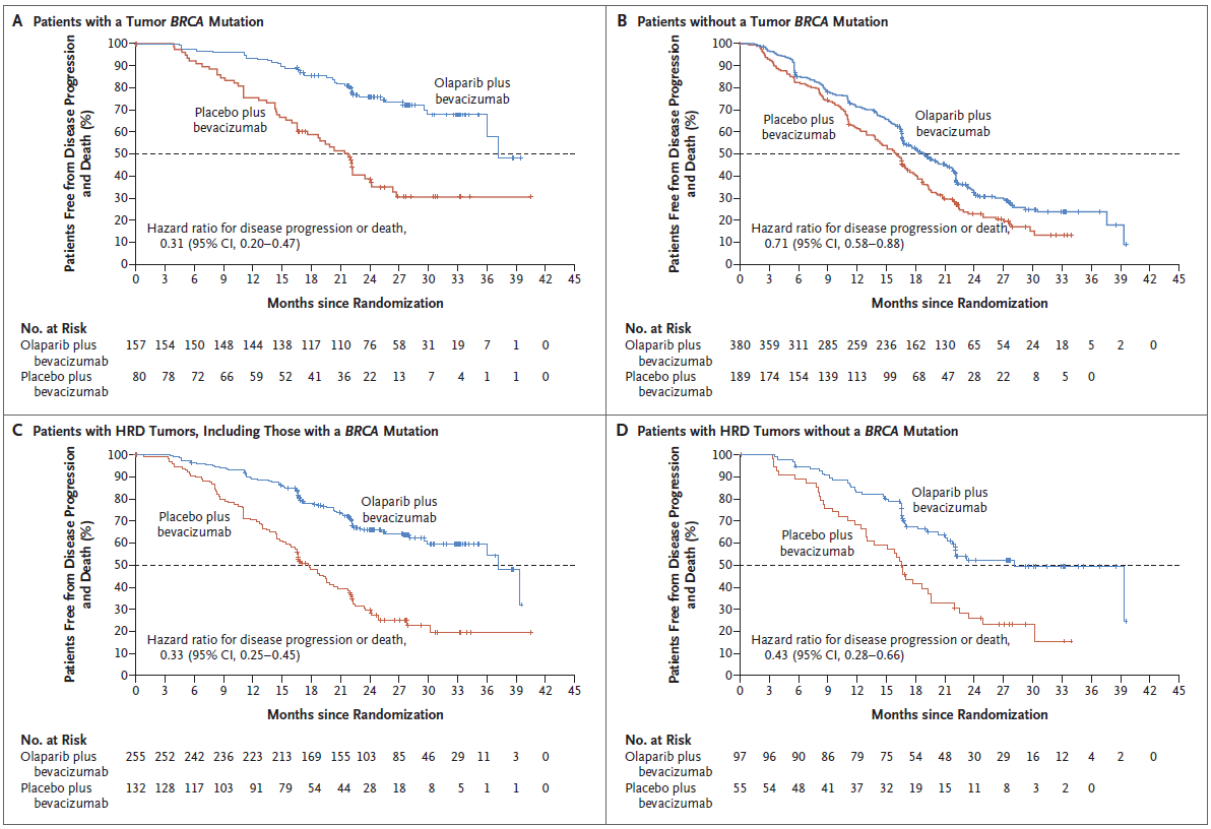
Figure 16 : Scores de déficience de recombinaison homologue sur différents types d'échantillons de tumeurs (d'après Abkevich et al. 2012). *Cercle bleu : Score de déficience de recombinaison homologue sur des données combinées de différentes cohortes avec des échantillons ayant un GPV dans BRCA1, BRCA2, ou RAD51C ou avec méthylation du promoteur de BRCA1 ou RAD51C. Cercles rouges : échantillons avec des gènes BRCA1, BRCA2, et RAD51C intacts.*

2. Application au versant somatique : essais cliniques

Un des intérêts majeurs de la détection des tumeurs HRD est son intérêt théranostique. En effet ces tumeurs sont connues pour être sensibles aux sels de platine mais aussi aux inhibiteurs de PARP. Ces

dernières années, les inhibiteurs de PARP n'ont cessé de démontrer leur intérêt dans les cancers de l'ovaire HRD et très récemment dans les cancers du sein HRD au travers de plusieurs essais cliniques :

- Dans l'essai SOLO-1, l'efficacité de l'olaparib en tant que traitement d'entretien a été évaluée chez des patientes atteintes d'un cancer de l'ovaire avancé, ou d'un cancer de l'ovaire séreux ou endométrioïde de haut grade, ou d'un cancer du péritoine primaire ou de cancer des trompes de Fallope (ou une combinaison de ceux-ci) porteuses d'une mutation constitutionnelle de *BRCA1*, ou *BRCA2*, ou les deux et ayant eu une réponse clinique complète ou partielle après chimiothérapie à base de platine. Après un suivi médian de 41 mois, le risque de progression de la maladie ou de décès était réduit de 70% avec l'olaparib contre placebo(97),
- En 2019 grâce à l'essai PAOLA-1, l'olaparib a démontré une diminution de la progression de la maladie ou du nombre de décès de 67% chez les patientes atteintes de tumeurs de l'ovaire HRD, *BRCA* mutées (survie médiane sans progression, 37,2 contre 17,7 mois), et de 57% chez les patientes atteintes de tumeurs HRD sans mutations *BRCA* – tumeurs *BRCAness* - (survie médiane sans progression, 28,1 contre 16,6 mois)(98)(**Figure 17**). Cette étude a démontré l'intérêt de détecter les tumeurs HRD au-delà des panels de gènes normalement testés. Des résultats similaires ont été obtenus avec l'essai PRIMA et le niraparib (**Tableau 2**)(99).



<u>Figure 17 :</u> Estimation Kaplan-Meier de la survie sans progression chez des patientes atteintes de cancer de l'ovaire séreux de haut grade, en fonction du statut tumoral BRCA et du statut de la déficience en recombinaison homologue (d'après Ray-Cocquard et al. 2019). A*. Parmi les patientes présentant une mutation BRCA tumorale le pourcentage de patientes sans progression de la maladie ni de décès à 24 mois était de 76% dans le groupe olaparib plus bevacizumab et de 39 % dans le groupe placebo plus bevacizumab. **B**. Parmi les patientes sans mutation BRCA tumorale le pourcentage de patientes sans progression de la maladie ni de décès à 24 mois était de 33% dans le groupe olaparib plus bevacizumab et de 23% dans le groupe placebo plus-bevacizumab. **C**. Parmi les patientes présentant des tumeurs HRD (définies par un score tumoral HRD de 42 ou plus) ou par une mutation BRCA tumorale, le pourcentage de patientes sans progression de la maladie ni de décès à 24 mois était de 66% dans le groupe olaparib plus bevacizumab et de 29% dans le groupe placebo plus bevacizumab. **D**. Parmi les patientes présentant des tumeurs HRD sans mutation BRCA le pourcentage de patientes sans progression de la maladie ni de décès à 24 mois était de 52% dans le groupe olaparib plus bevacizumab et de 26 % dans le groupe placebo plus bevacizumab.*

	Essai PRIMA			Essai PAOLA-1		
Sous-groupe	PFS (niraparib) en mois	PFS (placebo) en mois	Hazard Ratio (95%CI)	PFS (olaparib) en mois	PFS (placebo) en mois	Hazard Ratio (95%CI)
HRD/BRCAmt	22,1	10,9	0,40 (0,27-0,62)	37,2	17,7	0,33 (0,25-0,45)
HRD/BRCAwt	19,6	8,2	0,50 (0,31-0,83)	28,1	16,6	0,43 (0,28-0,66)
HRP	8,1	5,4	0,68 (0,49-0,94)	16,9	16	0,92 (0,72-1,17)

<u>Tableau 2 :</u> Comparaison des résultats entre l'essai PRIMA et PAOLA-1 évaluant respectivement l'efficacité du niraparib et de l'olaparib en entretien dans le cancer de l'ovaire séreux de haut grade selon le statut tumoral BRCA et le statut de la voie de la recombinaison homologue. *PFS : Progression Free Survival (survie sans progression) ; BRCAmt : BRCA muté ; BRCAwt : Sans mutation BRCA*

Enfin très récemment dans l'essai Olympia, l'olaparib a démontré une efficacité dans la survie sans progression chez des patientes atteintes de cancer du sein triple négatif, porteuses de GPV dans les gènes *BRCA1/2* après chimiothérapie néoadjuvante ou adjuvante. Avec un suivi médian de 2,5 ans, la survie sans maladie invasive à 3 ans était de 85,9 % dans le groupe olaparib et de 77,1 % dans le groupe placebo (IC à 95 %, 4,5 à 13,0), associé à une diminution du risque de maladie invasive ou de décès de 42% dans le groupe olaparib par rapport au groupe placebo (Hazard ratio de 0,58 (IC à 99,5%, 0,41 à 0,82; p <0,001))(100).

3. Application au versant constitutionnel : l'héritabilité manquante des tumeurs HRD

La signature HRD ne se retrouve pas uniquement dans les cancers du sein et de l'ovaire et s'étend à d'autres types de cancers. Des signatures HRD sont ainsi logiquement retrouvées dans les autres cancers du syndrome HBOC (cancer du pancréas, et de la prostate) mais aussi dans d'autres cancers en dehors de ce spectre (cancer du poumon, voies biliaires, etc…), alors que les gènes majeurs de la voie HR ne sont pas significativement impliqués dans leur oncogenèse(54)(**Figure 18**). Ce paradoxe reflète le fait qu'il existe peut-être d'autres gènes impliqués dans la physiopathologie de ces tumeurs HRD, non connus à ce jour. Citons par exemple que les tumeurs *IDH*, *FH* ou *SDH* mutées présenteraient une signature HRD, secondaire à une inhibition de KDM4B entraînant une hyperméthylation de l'histone H3K9me3, empêchant l'accès des protéines senseurs de la réparation (comme ATM) aux sites des cassures double brins(101).

Environ la moitié des cancers du sein et de l'ovaire ayant une signature HRD n'ont pas de cause génétique identifiée à ce jour, qu'elles soient constitutionnelles ou somatiques(54). Cependant, parmi les tumeurs HRD, environ 40% d'entre elles présentent une LOH seule dans les gènes HR sans 2nd altération identifiable (**Figure 18**).

Cette signature HRD peut donc servir à identifier de nouveaux gènes impliqués dans la voie HR et l'oncogenèse du syndrome HBOC(88), mais aussi à reclasser des VUS(102), ou à caractériser de nouveaux mécanismes mutationnels avec retentissement sur l'expression des gènes de la voie HR.

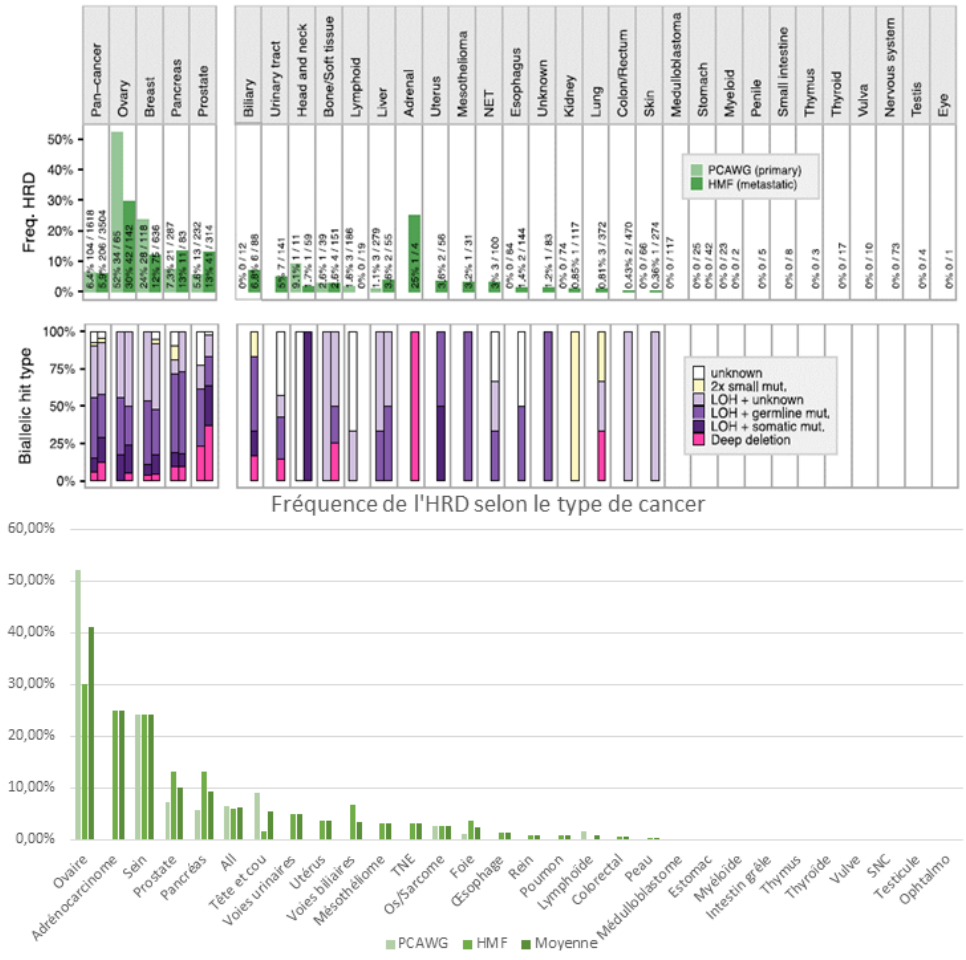


***Figure 18 :* Répartition en pourcentage de l'incidence de l'HRD par type de cancer et des causes de l'HRD chez les patients HRD positif selon CHORD-HRD (d'après Nguyen et al. 2020).** *Fréquences des signatures HRD selon le type de cancer issues de 3824 tumeurs métastatiques du Hartwig Medical Foundation (HMF) dataset et de 1618 tumeurs primitives issues du Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset. Combinaison probable d'événements bialléliques dans BRCA1/2, PALB2 ou RAD51C conduisant à une signature HRD. Fréquences de l'HRD triées par type de cancer.*

Cela démontre aussi que le séquençage tumoral associé à l'analyse constitutionnelle présente un intérêt étiologique en l'absence de mutation constitutionnelle dans les gènes HR permettant un rendement diagnostique plus important, tant au niveau étiologique que théranostique.

E. Stratégie de détection d'une signature HRD par bio-informatique et objectif de thèse

L'analyse de l'oncogenèse mammaire et ovarienne jusqu'à la signature HRD est rendue possible grâce à l'analyse en normal/tumeur et à la détection d'événements constitutionnels et somatiques. L'analyse constitutionnelle permet de détecter un événement initial de prédisposition comme la présence de GPV dans un gène de prédisposition mais aussi de collecter l'ensemble des polymorphismes propres à chaque individu. La connaissance du contexte constitutionnel permet de soustraire de l'analyse tumorale toutes les mutations qui ne sont pas propres à la tumeur mais aussi de détecter les seconds événements de grande taille apparus secondairement dans le paysage somatique de la tumeur (voir partie II.A.2). La connaissance des mutations propres à la tumeur permet par la suite une analyse fine des signatures mutationnelles et permet de dégager ou non une signature HRD, voire une autre (MMR, tabac, UV) concluant l'enquête oncogénique (**Figure 19**).

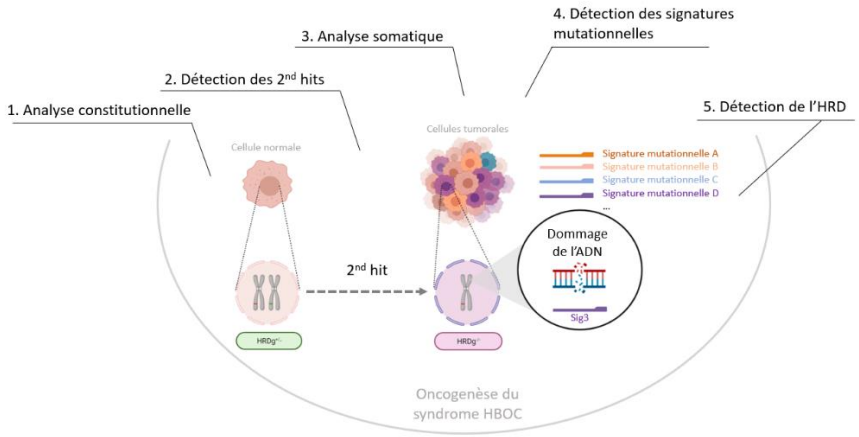


Figure 19 : Etude de l'oncogenèse dans le syndrome HBOC grâce à l'analyse en normal/tumeur en différentes étapes.

Objectif :

Extrapolation d'un pipeline bio-informatique de détection des tumeurs HRD par détection des signatures mutationnelles et des cicatrices génomiques sur une cohorte de patientes atteintes de cancers de l'ovaire en tissus FFPE, en analyse normal/tumeur par séquençages en panel de gènes, en situation de routine.

II. Méthodes

A. Développement d'un pipeline bio-informatique de détection d'une signature moléculaire HRD

1. Arsenal des outils bio-informatiques

En association à l'analyse en normal/tumeur, l'avènement de la bio-informatique ces dernières années a permis de mettre au point différents outils performants qui, assemblés sous la forme d'un « pipeline bio-informatique », permettent de détecter un large éventail d'événements impliqués dans l'oncogenèse mammaire et ovarienne (**Figure 20**). Ces outils sont utilisables sur différents environnements :

- En interface par ligne de commande, comme GATK ou R,
- Sur une interface mixte (ligne de commande, interface visuelle), comme les logiciels RStudio ou MATLAB,
- Sur des serveurs en ligne, comme GenePattern (*webtools*).

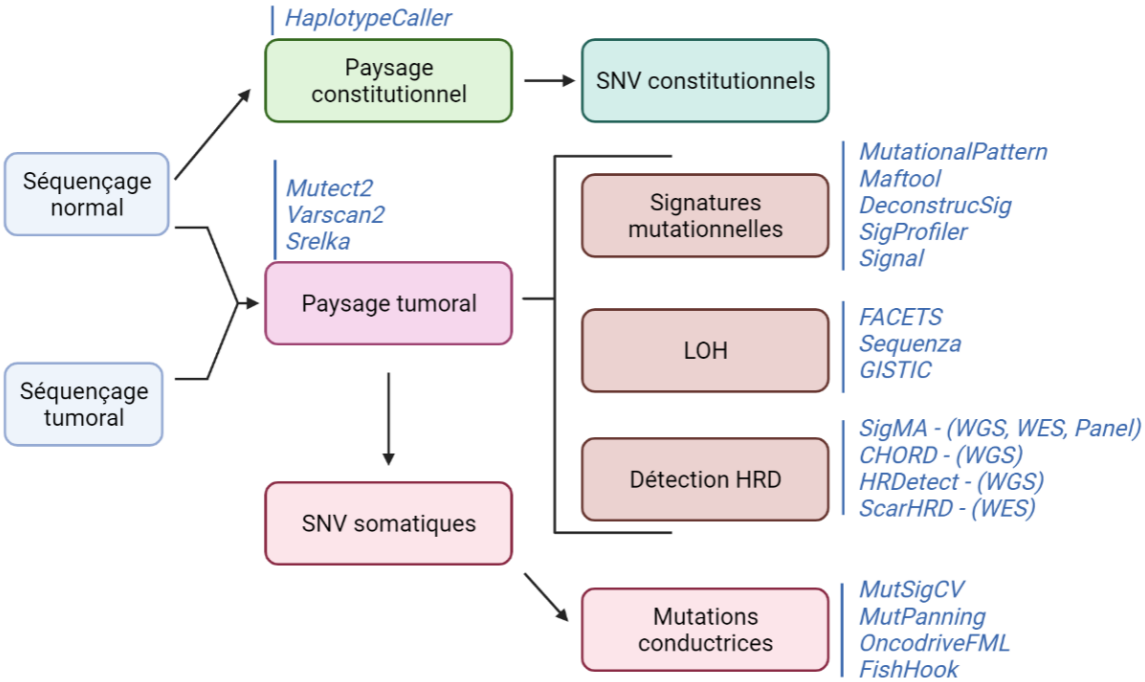


***Figure 20 :** Liste non exhaustive des outils bio-informatiques sous forme de pipeline et selon leur type de détection.*

a) Détection des événements au niveau constitutionnel

HaplotypeCaller de GATK(82)

HaplotypeCaller, est un *variant caller* à orientation constitutionnelle, un outil permettant la détection automatisée des variants (SNVs, et Indels de petites tailles) à partir de fichier BAM. L'appel de variants d'HaplotypeCaller se fait via le réassemblage local d'haplotypes basé sur une méthode dite « probabilistique » qui utilise des modèles statistiques pour estimer la probabilité de chaque génotype possible. Brièvement, à chaque fois que le programme rencontre une région présentant des variants, HaplotypeCaller rejette les informations du génome de référence existantes et réassemble complètement les reads dans cette région. D'autres outils existent, comme Varscan qui est basé sur une méthode dite « heuristique » utilisant des seuils comme la fréquence allélique, la couverture en read, ou un score de qualité pour valider ou non les variants.

Les avantages d'HaplotypeCaller consistent en :

- Une utilisation facilitée par une documentation simple et précise fournie par le Broad Institute,
- Une bonne sensibilité à détecter les indels.

b) Détection des événements somatiques

(1) Détection des SNVs et Indels

Mutect2 de GATK(103)

Mutect2 est un *variant caller* de variants somatiques basé sur la même méthode probabilistique qu'HaplotypeCaller, et disponible sur GATK. En y associant les données de séquençage d'un échantillon constitutionnel dit « normal » issu du même individu (analyse en pair normal/tumeur ou matché (voir partie II.A.2)), Mutect2 est conçu pour appeler uniquement les variants somatiques. Mutect2 permet de ne pas appeler les variants fréquents et/ou rares qui sont présents dans la lignée constitutionnelle sur la base de la séquence normale appariée. Cette opération est effectuée à un stade précoce de l'analyse afin d'éviter de dépenser des ressources informatiques sur les événements de la lignée

constitutionnelle. Si le statut constitutionnel du variant est limite, alors Mutect2 émettra le variant pour un filtrage ultérieur et une révision manuelle.

Pour augmenter sa spécificité dans l'appel de variants, Mutect2 utilise un « panel de normaux » (PoN pour *Panel of Normal*). Le PoN est un fichier VCF créé à partir d'au moins 40 échantillons normaux (dans ce contexte, "normal" signifie « dérivé de tissus sains qui sont supposés ne pas avoir d'altérations somatiques ») capturés et séquencés selon le même protocole que les échantillons tests. Son objectif principal est de capturer et lister les artefacts techniques récurrents dans un fichier VCF afin d'améliorer les résultats de l'analyse et les retirer de l'interprétation.

Mutect2 possède également un filtre interne de biais de brins permettant de répertorier les altérations sur chaque orientation de brins (*reverse* et *forward*) puis de les compter dans l'échantillon normal et l'échantillon tumoral. Ce filtre est particulièrement utile pour retirer de l'analyse les artéfacts secondaires au FFPE, et les possibles erreurs d'échantillonnages entre les échantillons normaux et tumoraux.

De plus, afin de diminuer le taux de faux-positif, Mutect2 utilise une base de données externe de gnomAD, contenant une liste exhaustive des mutations retrouvées en population générale.

D'autres *variant callers* somatiques peuvent être utilisés comme Varscan ou Strelka, basés sur des modèles mathématiques différents(104). Un des atouts de Mutect2 par rapport aux autres *variant callers* est sa grande spécificité dans l'appel de variants. Dans une étude de marché des *variant callers*, bien que Mutect2 ait proposé moins de variants candidats par rapport à Strelka2 ou Varscan, il avait une excellente spécificité(105).

Mutect2 et Strelka2 ont des performances similaires et font partie des *variant callers* les plus fiables pour l'appel de SNV(104). L'avantage de Strelka2 est une détection et une précision plus élevées que Mutect2 lorsque la fréquence de mutation est élevée (≥20%), alors que Mutect2 est plus performant lorsque la fréquence de mutation est inférieure à 10%, ce qui est particulièrement utile pour détecter des clones tumoraux de faible fréquence.

Ainsi, Mutect2 s'impose comme un *variant caller* de qualité avec une bonne spécificité, un atout pour détecter des variants issus de tumeurs FFPE pourvoyeuses d'artéfacts.

Une stratégie récente est l'utilisation combinée de multiples *variant callers* pour augmenter la spécificité des variants somatiques retrouvés(106,107). Par cette méthode, le taux de détection de variants est plus performant, tout en gardant une bonne spécificité. Les variants non détectés par un outil, pouvant être récupérés par un autre.

(2) Détection des mutations conductrices

Il existe deux types de mutations dans l'oncogenèse : les mutations conductrices (*driver*) de l'oncogenèse, et les mutations opportunes ou passagères (*passenger*). Les mutations conductrices sont impliquées dans l'oncogenèse alors que les mutations passagères sont présentes dans le génome tumoral mais n'ont, à priori, aucune implication pathogénique (**Figure 21**). Les mutations conductrices peuvent être de type gain de fonction dans un oncogène ou de type perte de fonction dans un gène suppresseur de tumeur.

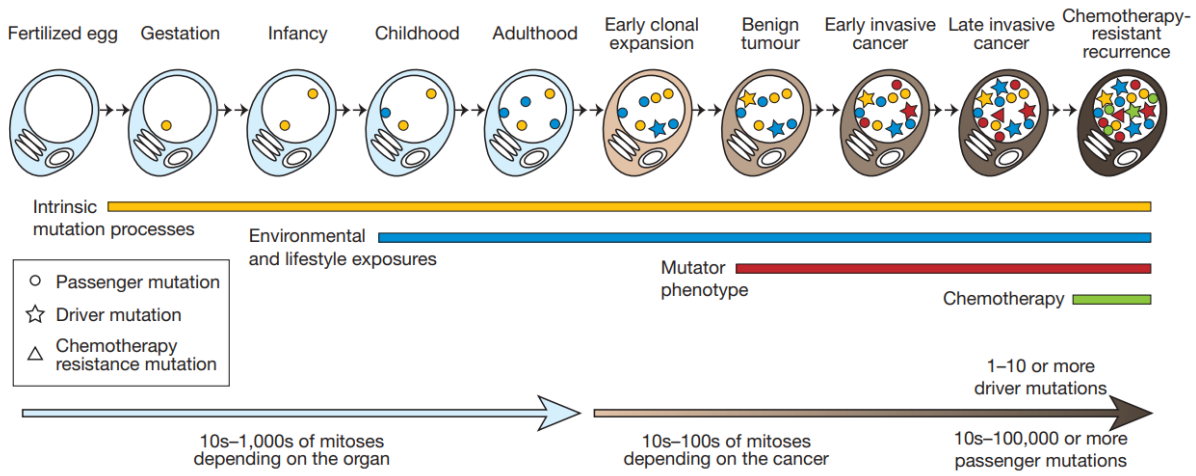


Figure 21 : Etapes de l'oncogenèse (d'après Stratton et al. 2009). *Les mutations passagères (représentées sous la forme d'un rond) sont présentes dès la gestation, et s'accumulent tout le long de la vie de l'individu, ainsi que dans la tumeur. Les mutations conductrices (représentées sous la forme d'une étoile) apparaissent en même temps qu'émerge un clone tumoral, puis s'accumulent en provoquant une instabilité génétique (Mutator phenotype).*

Détecter ces mutations présente plusieurs intérêts : mieux comprendre les mécanismes de l'oncogenèse, dresser le paysage génétique de la tumeur, identifier des cibles thérapeutiques potentielles, et rendre éligible un patient à une thérapie ciblée quand celle-ci est disponible.

Le principal enjeu de la détection des mutations somatiques est de caractériser un gène qui n'est pas muté par hasard. Cette détection est dépendante du fond génétique de la tumeur et varie en raison de facteurs propres à la tumeur et de facteurs génomiques basés sur la position d'une mutation. Les facteurs propres à la tumeur sont : le taux de mutation et le spectre global des mutations (comme le pourcentage de substitutions de type transition ou transversion et/ou des non-sens). Les facteurs basés sur la position génomique comprennent : les niveaux d'expression des gènes, le temps de réplication de l'ADN, l'état de la chromatine basé sur des données de HiC.

De nombreux outils sont disponibles pour détecter ces mutations conductrices, citons ici quelques outils utilisés dans ce travail :

- **MutSigCV**(108)**,** est un outil développé par le Broad Institute et utilisable en accès libre sous forme de module sur le site GenePattern. MutSigCV détecte les mutations conductrices en utilisant la fréquence des mutations retrouvées, les spectres de mutation spécifiques aux patients, ainsi que les taux de mutation spécifiques aux gènes, en y intégrant les niveaux d'expression et les temps de réplication,
- **MutPanning**(109), est un outil développé par le Broad Institute et utilisable en accès libre sous forme de module sur le site GenePattern. MutPanning se base sur le fait que la plupart des mutations passagères se produisent dans des contextes nucléotidiques "inhabituels" caractéristiques qui reflètent le processus mutationnel actif dans une tumeur donnée,
- **OncodriveFML**(110), est un *webtool* en accès libre qui permet d'estimer un biais d'impact fonctionnel pour des mutations somatiques dans des régions codantes et non codantes, sur la base d'une simulation des processus mutationnels qui les affectent,

- **FishHook**(111)**,** est un package R qui permet une analyse statistique de la récurrence des mutations codantes et non codantes dans un tissu tumoral par le biais d'une modélisation linéaire généralisée de la densité des mutations somatiques et de leur hétérogénéité le long du génome.

Les résultats sont donnés sous forme d'un q-score, reflétant la probabilité de trouver une mutation conductrice par erreur. Le seuil de q-score inférieur à 0,05 a été retenu.

(3) Détection des CNA et des LOH

La détection des CNA et des LOH, qui sont à l'origine des *second hits*, est une étape importante pour la compréhension de l'oncogenèse de la majorité des syndromes de prédispositions aux cancers.

GATK-CNA(82)

GAKT CNA est l'algorithme de détection des CNA somatiques proposé par GATK. GATK-CNA utilise différents logiciels de GATK pour cette analyse. La détection des CNA se faisant le plus souvent par analyse en CGH-array en comparant sur une puce à ADN le nombre de copies d'une région génomique *a* d'un échantillon A à tester par rapport au nombre de copies d'une région génomique *b* à un échantillon B de référence. Les résultats sont ensuite exprimés en log2 ratio : $Log2(\frac{a}{b})$. GATK-CNA rend possible la détection des CNA par analyse NGS via la création d'un PoN similaire à celui de Mutect2, comme set de référence (**Figure 22**). Cependant, cet algorithme ne permet pas de détecter les pertes d'hétérozygotie par copie neutre.

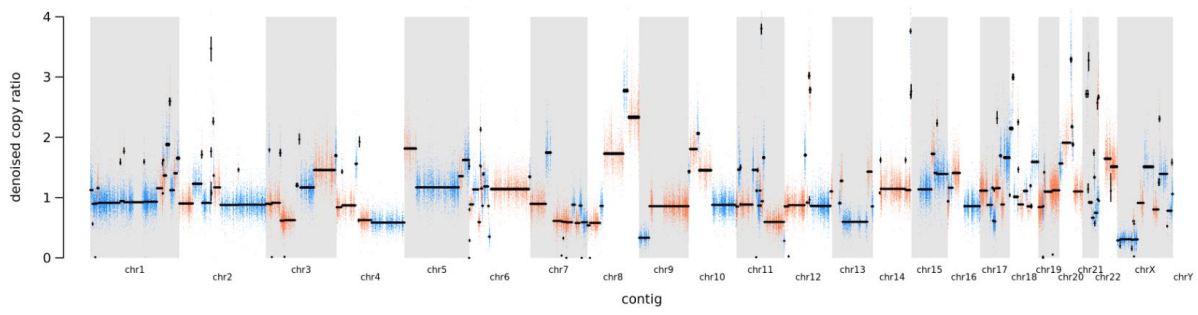


***Figure 22 :* Chromosome-plot d'un échantillon tumoral généré à l'aide GATK-CNA.** *Le nombre de copies de chaque segment est donné en log ratio. Les segments sont représentés par une barre noire horizontale, et*

verticale lorsqu'ils sont trop petits. Les alternances de couleurs bleu/rouge signifient une différence du nombre de copies entre chaque segment génomique analysé. Les segments conjoints sont de la même couleur.

Les résultats de GATK-CNA sont donnés sous forme de fichier SEG, exploitable pour des analyses complémentaires ou à des fins de visualisation par d'autres logiciels.

GISTIC2.0(112)

GISTIC2 est un outil développé par le Broad Institute et utilisable en accès libre sous forme de module sur le site GenePattern reconnu pour la détection et la visualisation des CNA. Le module GISTIC permet d'identifier les régions du génome qui sont significativement amplifiées ou délétées dans un ensemble d'échantillons (G-score et q-score), et de dresser la liste des gènes retrouvés dans ces régions. Les résultats sont donnés sous forme de fichier GISTIC exploitable pour des analyses complémentaires ou à des fins de visualisation par d'autres logiciels.

FACETS(113)

FACETS est un package R d'analyse du nombre de copies spécifiques d'allèles permettant la détection des délétions homozygotes/hétérozygotes, des pertes d'hétérozygotie par copie neutre, et des gains/amplifications. FACETS est un outil autonome qui inclut le post-traitement des fichiers BAM, une segmentation du compte des reads totaux et spécifiques d'allèles, et une détection du nombre intégral de copies ajusté sur la pureté tumorale, la ploïdie et l'hétérogénéité clonale.

Brièvement, FACETS utilise une approche de segmentation conjointe non paramétrique basée sur une statistique T2 de Hotelling en combinant directement les comptes des reads totaux et spécifiques d'allèles via une analyse en normal/tumeur. FACETS utilise une approche basée sur les SNP, car le déséquilibre allélique ne peut être mesuré que sur les sites hétérozygotes. Ainsi le nombre de reads tumoraux au niveau de chaque segment homozygote par rapport au nombre de reads constitutionnels permet d'évaluer une altération du nombre de copies et ce, tout le long du génome. Cette donnée est exprimée en Log-ratio. Seule, cette analyse ne permet pas d'identifier des pertes d'hétérozygotie par copie neutre, car le nombre total de reads tumoraux serait le même que le nombre total de read

constitutionnels, et donc un Log-ratio à 0 (**Figure 23**). C'est pourquoi FACETS utilise une liste de SNP connus afin de pouvoir évaluer le nombre de reads porteurs de l'allèle majeur (dit allèle A) le plus fréquent en population générale et le nombre de reads porteurs de l'allèle mineur (dit allèle B) ou polymorphe, au niveau de sites hétérozygotes connus à la fois dans le tissu tumoral et le tissu sain. Cette donnée est exprimée en Log-OddsRatio (**Figure 23**).

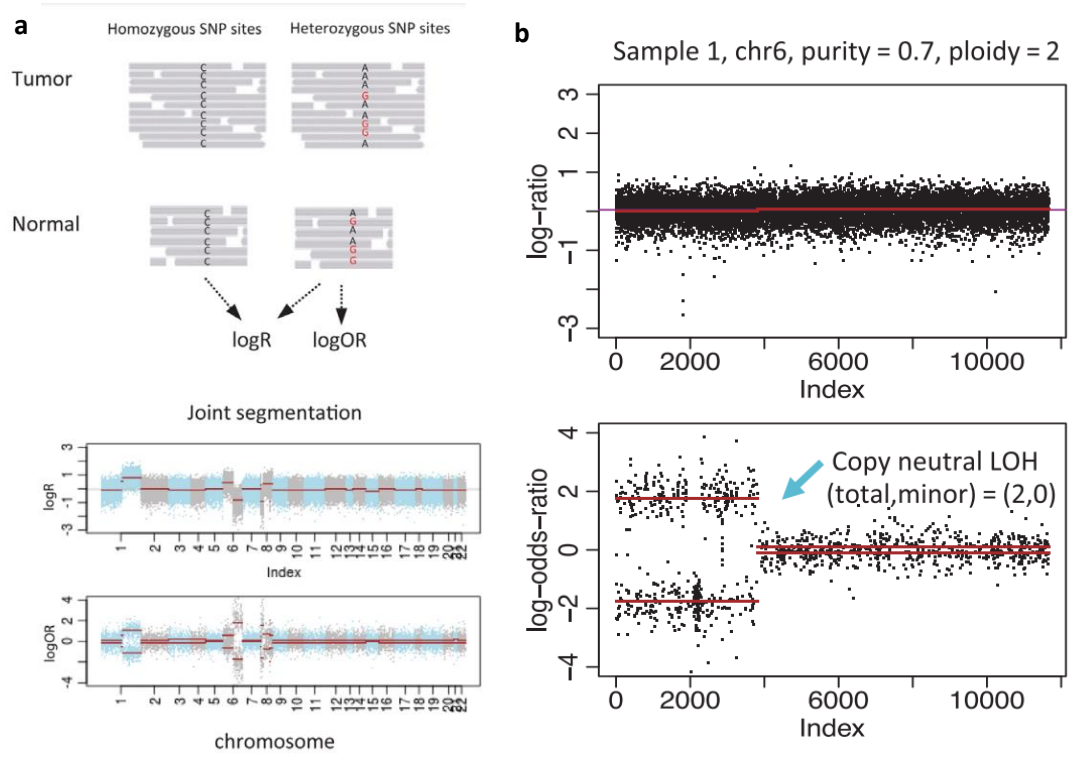


Figure 23 : Principe de détection des CNA et des LOH selon FACETS. a) *Analyse du nombre d'allèles de référence et de variants pour les sites SNP à partir des fichiers BAM normal/tumeur. Tous les sites SNP contribuent au log-ratio de copie total (logR), et les sites hétérozygotes contribuent au logOR allélique. Segmentation conjointe logR et logOR et détection des régions aberrantes en nombre de copies du génome. La segmentation conjointe permet d'identifier les événements par perte d'hétérozygotie par copie neutre (Copy neutral LOH).* ***b)*** *Le panneau supérieur montre le log-ratio du nombre total de reads dans la tumeur par rapport à celui de la séquence normale le long des positions génomiques sur le chromosome 6, ici il ne semble pas y avoir de gain ni perte. Le deuxième panneau montre le log-odds-ratio allélique du nombre de reads de l'allèle variant dans la paire normal/ tumeur, révélant un événement LOH par copie neutre sur le chromosome 6p grâce à la présence de SNP à l'état homozygote dans la tumeur.*

Grâce à cette méthode, si une perte d'hétérozygotie par copie neutre se produit sur un segment génomique, un des deux allèles sera sur-représenté par rapport à l'autre, et le LogOR allélique sera en faveur d'une délétion ou d'une duplication selon l'allèle. Du LogOR de chaque segment est déduit la ploïdie, sous la forme d'un logRo. Le calcul du nombre intégral de copies observées le long du génome

(de chaque segment conjoint) est obtenu par la formule exp(logRc – logRo), où LogRc est le LogR de chaque segment conjoint. FACETS scanne le génome par segment toutes les 150-200pb, et lorsqu'ils ont un LogR identique, les rassemble sous forme de segments conjoints. Les résultats de FACETS sont donnés sous forme de fichier VCF, exploitable pour des analyses complémentaires par d'autres logiciels.

Sequenza(114)

Sequenza est un logiciel de détection et visualisation de la LOH en deux parties : un outil de prétraitement écrit en language Python et un package R pour la visualisation. Sequenza est basé sur un modèle probabiliste de détection de LOH par segmentation, pour une analyse en normal/tumeur. Presque similaire à FACETS, Sequenza prend en compte dans son analyse le rapport de la profondeur de séquençage moyen (tumeur par rapport au contingent normal) et la fréquence de l'allèle mineur pour chaque segment (la plus petite des deux fractions alléliques mesurée aux positions hétérozygotes constitutionnelles). Les autres analyses comprennent aussi le calcul de la ploïdie et la cellularité tumorale, le nombre de copies intégrales pour chaque segment analysé et le nombre de copies de l'allèle mineur. Les résultas de Séquenza sont donnés sous forme de fichier SEQZ.GZ, exploitables pour des analyses complémentaires par d'autres logiciels.

c) Détection des signatures mutationnelles

Il existe de nombreux outils de détection des signatures mutationnelles, le plus souvent en accès libre sur différentes plateformes. On peut citer SigProfiler disponible sur MATLAB, DeconstructSigs, MutationalPattern, ou Maftools disponibles sur R, MuSiCa, Mutalisk ou Signal disponible en ligne sous forme de *webtool*. Tous ont des caractéristiques propres et ne servent pas les mêmes usages ou n'utilisent pas le même type de fichier d'entrée. Pour ce projet, nous avons utilisé des outils disponibles sur R : MutationalPattern, DeconstructSigs et Maftools, utilisant des fichiers VCF ou MAF, ainsi que Signal, utilisant des fichiers VCF.

MutationalPattern(115)

MutationalPattern est un package R de détection des signatures mutationnelles. Il utilise des fichiers VCF en entrée. MutationalPattern fournit un ensemble complet de fonctions flexibles qui permettent d'évaluer et de visualiser facilement une multitude de modèles mutationnels via l'analyse de substitution de base. Le package fonctionne avec des variants nucléotidiques simples (SNV), des insertions et des délétions (Indels), des substitutions à doubles bases (DBS) et des substitutions de bases multiples plus importantes (MBS). Le package fournit des fonctionnalités pour extraire des signatures mutationnelles *de novo* par factorisation par matrice non négative (NMF) et déterminer la contribution des signatures mutationnelles précédemment identifiées comme les signatures COSMIC au niveau d'un seul échantillon (**Figure 24**).

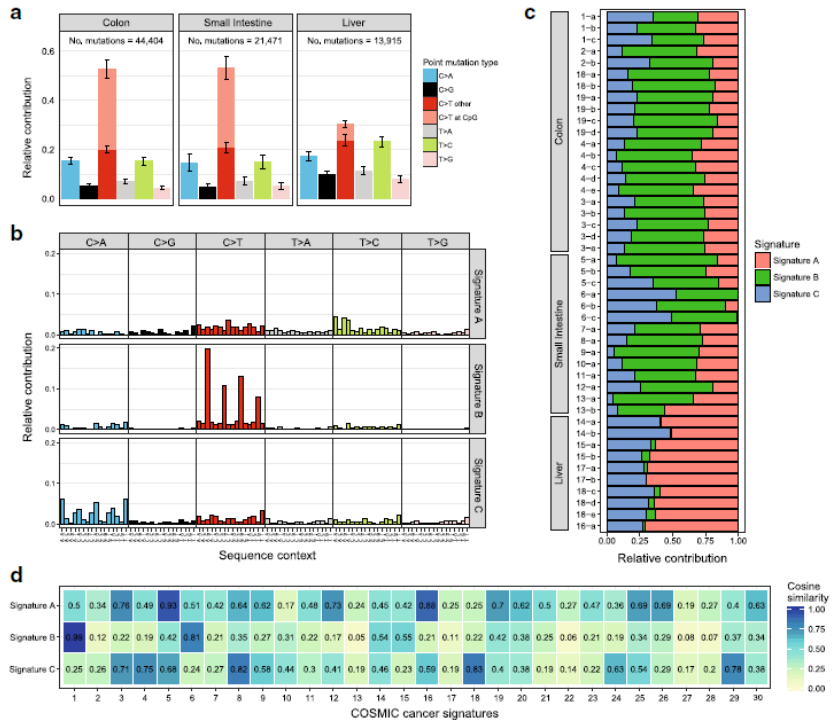


Figure 24 : Caractéristiques des mutations somatiques acquises sur des cellules souches humaines (CSH) de différents tissus à l'aide de MutationalPattern. a) *Contribution relative des types de mutation pour chaque type de tissu. Les barres représentent la contribution relative moyenne de chaque type de mutation par type de tissu et les barres d'erreur indiquent l'écart type. Le nombre total de mutations somatiques ponctuelles par tissu est indiqué.* ***b)*** *Contribution relative de chaque changement trinucléotidique dans les trois signatures mutationnelles « de novo » identifiées par analyse NMF dans les paysages mutationnels somatique des CSH.* ***c)*** *Contribution relative de chaque signature mutationnelle « de novo » pour chaque échantillon.* ***d)*** *Heatmap montrant la similarité cosinus des signatures mutationnelles « de novo » avec le catalogue des signatures COSMIC. Ici la signature B semble correspondre à la signature COSMIC 1 lié à l'âge.*

DeconstructSig(116)

DeconstructSig est un package R de détection des signatures mutationnelles. Il utilise des fichiers VCF en entrée. DeconstructSigs vise à déterminer la contribution des processus mutationnels connus (signature COSMIC) dans un échantillon tumoral. En utilisant DeconstructSigs, on peut déterminer le poids de chaque signature mutationnelle contribuant à un échantillon tumoral individuel, tracer le profil mutationnel reconstruit (en utilisant les poids calculés) et le comparer à l'échantillon d'entrée original.

Module de Signature Mutationnelle de Maftools(117)

Maftools est un package R de visualisation de données, utilisant des fichiers MAF en entrée. Les fichiers MAF (pour *Mutation Annotation Format*) diffèrent des fichiers VCF car ils possèdent des informations quant aux conséquences et caractéristiques des mutations retrouvées. Les fichiers VCF ont donc été convertis en fichier MAF en utilisant l'outil Vcf2maf(118). Maftools permet d'extraire des signatures mutationnelles *de novo* et de déterminer la contribution des signatures mutationnelles précédemment identifiées comme les signatures COSMIC au niveau d'un seul échantillon.

d) Caractérisation de l'HRD par les techniques NGS

La majorité des outils de détection de l'HRD ont été conçus pour des études en WGS ou WES, en utilisant des tissus tumoraux frais congelés. Il n'existe pas, à ce jour, d'outil de détection des signatures mutationnelles conçu spécifiquement pour des tissus FFPE, bien que ces programmes prennent en compte les artéfacts de séquençage dans leur analyse finale (**Tableau 3**).

Techniques de séquençage	Outils de détection HRD	Type d'analyse HRD
WGS	HRDetect	Signature nucléotidique
WGS/WES	Signature3	Signature nucléotidique
sWGS	ShallowHRD	Cicatrice génomique
WES	ScarHRD	Cicatrice génomique
WES/Panel	SigMA	Signature nucléotidique
snp-array	LST/TAI/LOH	Cicatrice génomique

***Tableau 3 :* Liste des différents outils bio-informatiques et fonctionnels de détection de l'HRD, selon le type d'analyse de séquençage**

SigMA(119)

SigMA est un package R d'analyse de signature optimisé pour la détection des signatures mutationnelles associée à l'HRD pour des analyses en panels de captures (panel de 410 gènes MSK-IMPACT), de WES et de WGS et adaptés aux sous-types tumoraux. SigMA peut utiliser en entrée des fichiers VCF ou MAF.

Les profils de signatures mutationnelles extraits d'analyses en WGS ont été simulés/corrélés à des profils en WES et de panel via *machine learning*. Grâce à une analyse multivariée, SigMA donne un score final pour la présence de la signature SBS3 en combinant la vraisemblance avec la similarité cosinus et l'exposition de la signature SBS3 dans différents sous-types tumoraux, à l'aide de données issues de WGS des consortiums ICGC et TCGA (**Figure 25**). La probabilité que l'échantillon appartienne à chaque sous-type de tumeur est calculée, et la probabilité de la signature SBS3 a été définie comme la somme des probabilités de tous les sous-types de tumeurs exprimant une signature SBS3.

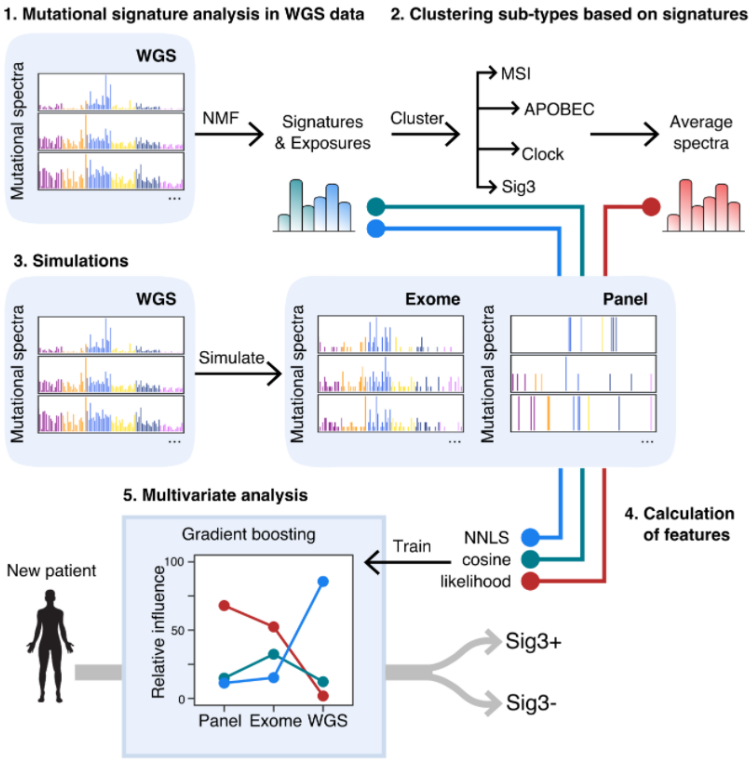


Figure 25 : Les étapes de SigMA*. L'analyse de signature est effectuée avec des données WGS et les expositions de signature de chaque échantillon sont déterminées. Des exomes et des panels simulés sont générés par sous-échantillonnage à partir des données WGS. Pour les données simulées, plusieurs statistiques (comme la similarité cosinus, la vraisemblance) sont calculées. Un classificateur à apprentissage automatique est entraîné en utilisant ces caractéristiques. Pour les panels, la mesure de vraisemblance reçoit plus de poids dans la prédiction.*

ScarHRD(120)

ScarHRD un package R de détection de l'HRD via l'étude des cicatrices génomiques conçu pour WGS et WES. ScarHRD rend possible l'analyse par NGS des sous-scores LOH, LST, et TAI (initialement détectés sur des analyses en SNP-array(121)) avec une très bonne corrélation ($R^2 = 0,75$, $p < 2,2e^{-16}$). Ces sous-scores sont ensuite combinés en score HRD (HRD Sum Score). Comme les analyses via SNP-array, un score ≥ 42 (seuil à partir duquel 95% des tumeurs déficientes pour les protéines BRCA1/2 sont considérées comme HRD via les études de SNP-array) permet de classer les tumeurs comme étant HRD(122).

2. L'analyse en normal/tumeur

Le séquençage en paires d'échantillons dit normal/tumeur (« analyse en normal/tumeur ») est un processus de détection des mutations somatiques plus précis que le séquençage de tumeur seul grâce à la prise en compte du contingent constitutionnel dans l'analyse. Grâce au séquençage simultané du tissu sain et du tissu tumoral appartenant à un même individu, nous obtenons le fond génétique constitutionnel et tumoral. Ensuite, par bio-informatique, nous pouvons soustraire du contingent tumoral le fond génétique constitutionnel pour ne garder que les mutations somatiques. Ainsi, le paysage mutationnel de la tumeur peut pleinement être analysé, sans prendre en compte des mutations constitutionnelles ou des polymorphismes fréquents, réduisant la fenêtre d'erreur d'interprétation (**Figure 26**).

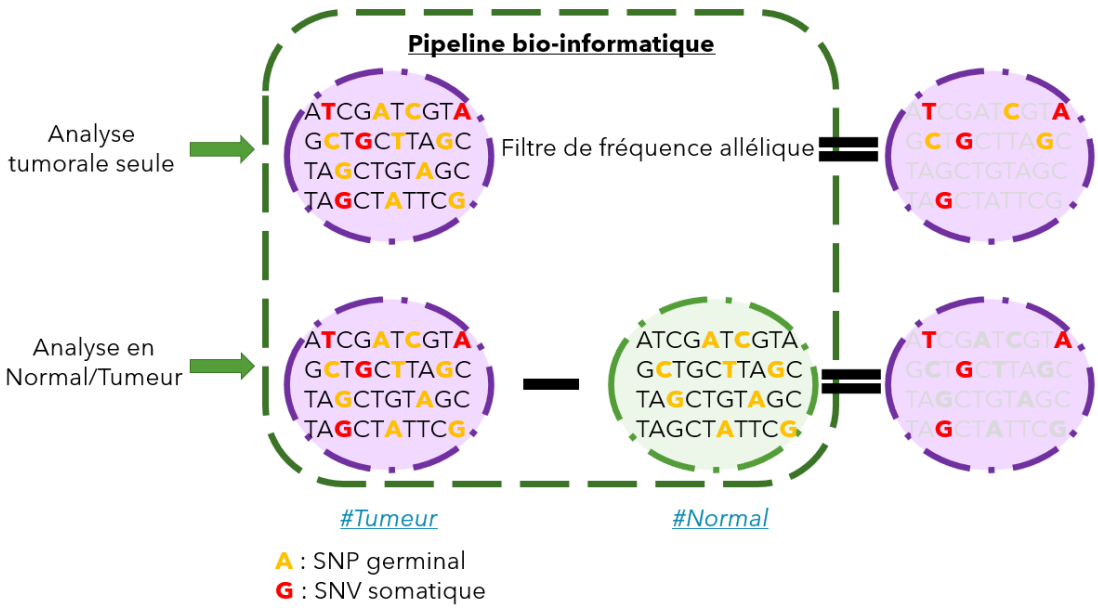


***Figure 26 :* Schéma des différentes analyses du génome tumoral.** *L'analyse tumorale seule nécessite d'utiliser différents filtres comme la fréquence du variant dans l'échantillon analysé ou l'utilisation de bases de données externes. L'analyse en normal/tumeur permet de s'abstenir de l'usage de filtre de fréquence car les variants présents dans les deux contingents sont automatiquement soustraits. SNP : Single Nucleotide Polymorphism, SNV : Single Nucleotide Variant*

Initialement développé pour mieux comprendre les processus oncogéniques somatiques de certaines tumeurs (123) ainsi que dans l'évolution clonale métastatique(124), ce procédé a été utilisé dans la détection des signatures mutationnelles(92). Bien que de très bonne qualité, ces analyses sont

malheureusement difficiles à transposer en routine car elles nécessitent l'utilisation de tissus frais congelés difficiles à obtenir, coûteux en stockage et difficiles à réutiliser pour des analyses ultérieures.

En situation de routine et dans les laboratoires d'anatomopathologie, les tumeurs sont le plus souvent enrobées et conservées dans du formaldéhyde fixé en paraffine dit tissu « FFPE » *(Formalin Fixed Paraffin Embedded*). Malheureusement, en biologie moléculaire, le FFPE est pourvoyeur d'altérations de l'ADN et générateur d'artefacts de séquençage aléatoires et non prévisibles(125). La fixation de tissu avec le formaldéhyde induit des dommages à l'ADN comme la désamination hydrolytique de la cytosine en uracile. Il en résulte des artefacts de séquençage C>T/G>A non reproductibles, qui sont observés après l'étape d'amplification en NGS(125,126). Ces variants artéfactuels ne sont pas présents sur toutes les molécules d'ADN, se présentant sous forme de biais de brins et dont la fréquence allélique dans la tumeur avoisine souvent moins de 25%(125). C'est pourquoi de nombreux outils bio-informatiques ont été développés ces dernières années afin de prendre en compte ces artéfacts et les retirer de l'analyse somatique comme le fait Mutect2(103). Il existe aussi d'autres procédés comme l'application manuelle de filtres basée sur la fréquence allélique des variant (VaF), le nombre de reads porteurs de variants somatiques ou la combinaison de multiples *variant callers* pour augmenter la spécificité des variants somatiques retrouvés(106,107).

Cependant, bien que de plus en plus performants, ces outils ne suppriment pas tous les artéfacts, et filtrer sur la VaF augmente le risque de retirer l'analyse des clones tumoraux de fréquence faible inférieure à 20%, et donc de véritables mutations somatiques. Actuellement il n'y a pas de recommandation de filtrage des variants artéfactuels sur un seuil de VaF optimal, même si un seuil entre 2-20% est communément appliqué. Il est aussi d'usage de filtrer le nombre minimal de reads porteurs de véritables variants somatiques entre 0 et 20(102,125,127).

Enfin, l'utilisation de prélèvement sanguin nécessaire à l'analyse en normal/tumeur est une tâche malaisée, car les patients ne sont pas prélevés en parallèle de leur biopsie tumorale. En France notamment, les analyses génétiques constitutionnelles étant entourées d'un cadre légal strict, il est

nécessaire de récupérer les consentements d'autorisation de tests génétiques. Ce type de procédé est difficilement standardisable à l'heure actuelle et limité à quelques équipes ou centres de recherche. C'est pourquoi, plusieurs équipes ont développé des pipelines bio-informatiques afin de permettre l'utilisation de tissus sains FFPE comme contingent normal avec une bonne corrélation tissus sanguin/tissus sains FFPE(127).

3. Choix des outils bio-informatiques et mise en place du pipeline

Un premier pipeline de détection d'une signature HRD a été développé pour une analyse en normal/tumeur par WES contenant les divers outils et paramètres ci-dessous(128)(**Figure 27**). Ce pipeline a ensuite été légèrement modifié pour utilisation en routine (voir partie III.B).

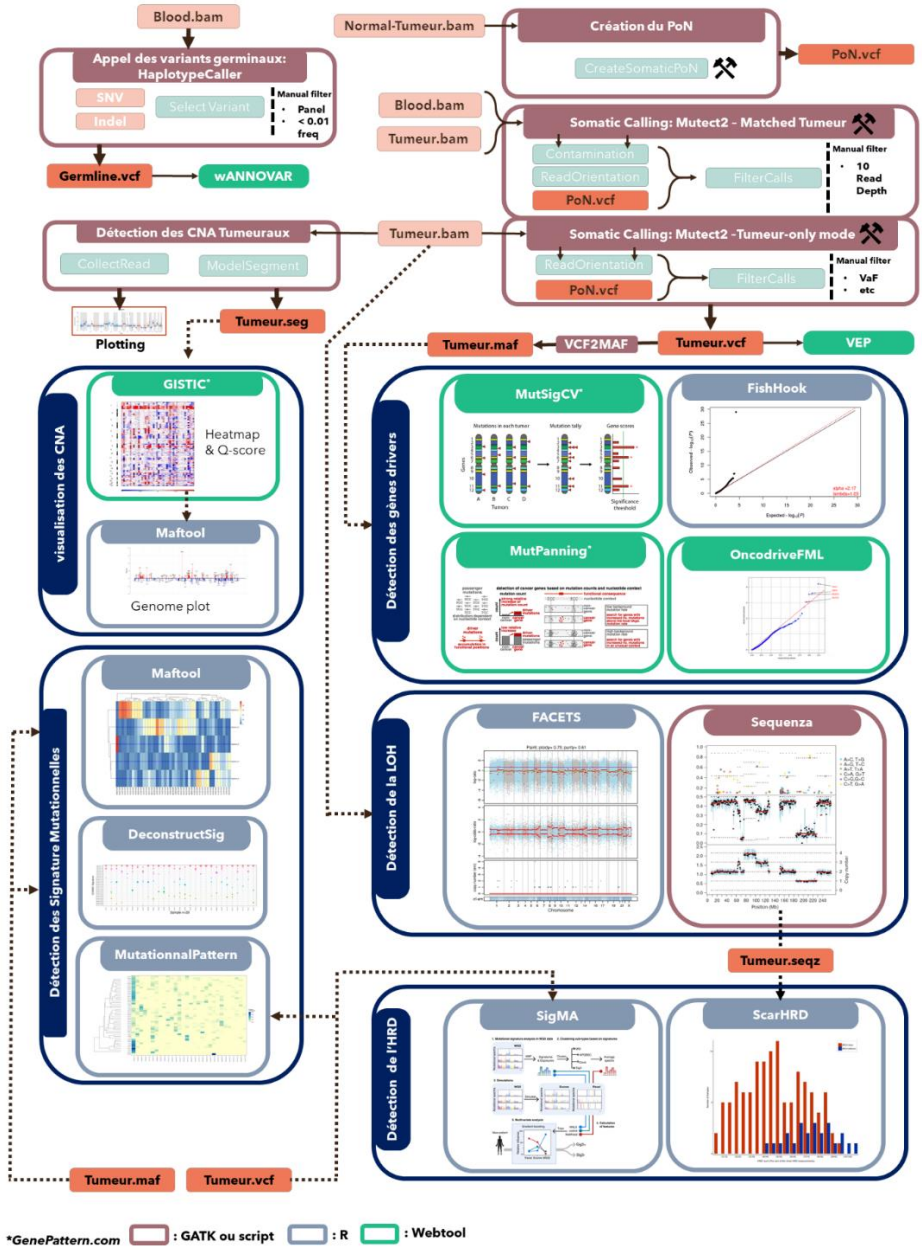


Figure 27 : Vignette du pipeline bio-informatique initial développé pour la détection de l'HRD

a) Appel et filtrage des variants constitutionnels

BWA (v. 0.5.9) et GATK sont utilisés pour aligner les reads séquencés sur le génome de référence (UCSC hg19) et pour effectuer un réalignement local des reads autour des petites insertions et délétions. Les reads dupliqués par PCR sont marqués à l'aide de Picard (https://github.com/broadinstitute/picard). Les variants constitutionnels sont appelés individuellement pour chaque échantillon en utilisant HaplotypeCaller de GATK(82) et annotés par wAnnovar(129) qui est un *webtool* en libre accès d'annotation des variants. Nous avons choisi HaplotypeCaller pour son utilisation facile et son guide de bonnes pratiques, sans besoin de connaissances approfondies en bio-informatique pour son

installation et interprétation. À partir de la liste de variants constitutionnels générée, les variants sont éliminés par filtrage s'ils remplissent l'un des critères de qualité suivants : (i) position génomique du variant couverte par <5 reads ; (ii) <2 reads soutenant un variant alternatif ; (iii) ; fraction allélique entre < 0,35 ou > 0,5 ; (iv) et qualité de mapping < 30.

b) Appel et filtrage des variants somatiques

Les variants somatiques sont appelés individuellement via le processus de *variant caller* en normal/tumeur de Mutect2 pour les SNV et les indels. Un PoN développé pour l'occasion est utilisé ainsi que les ressources de la population générale (gnomAD) conformément aux bonnes pratiques de GATK et du Broad Institute. Mutect2 étant un *variant caller* stringent, il nous a semblé intéressant de l'utiliser pour essayer de s'affranchir au maximum des variants faux positifs dus à des artéfacts FFPE, quitte à obtenir moins de variants de l'analyse. Les variants sont éliminés par filtrage s'ils (i) présentent un biais de brin, afin d'éliminer les artefacts causés par le processus de fixation au FFPE ; (ii) sont d'origine constitutionnelle ; (iii) sont dus à des artefacts techniques révélés par la comparaison avec le PoN, et (iv) ont une profondeur < 10 reads. Ce dernier seuil a été choisi arbitrairement pour s'affranchir de possibles artéfacts FFPE de faible fréquence. Les variants ont ensuite été annotés avec VEP, un *webtool* en libre accès(130). Le logiciel IGV a été utilisé pour vérifier la présence des variants sur les fichiers BAM.

c) Analyse du nombre de copies et détection de LOH somatiques

Le compte de couverture de reads à travers les zones de capture ont été collecté grâce à GATK-CNA. Un PoN distinct a été créé pour la variation du nombre de copies (CNV) afin de réduire le bruit de fond des données de couverture des reads et d'obtenir des ratios de copies de dé-bruitage par rapport au PoN. Les comptes des allèles de référence par rapport aux allèles alternatifs pour le déséquilibre allélique sont ensuite recueillis. GISTIC(112) est utilisé pour évaluer les amplifications et les délétions focales récurrentes dans les tumeurs. Nous avons utilisé FACETS(113) et Sequenza(114) pour déduire le nombre de copies intégrales de chaque paire d'allèles ainsi que la LOH. Les délétions hémizygotes sont définies comme la perte d'un allèle dans la tumeur. Les délétions homozygotes sont définies

comme la perte des deux allèles dans la tumeur. L'amplification a été définie comme un statut de nombre de copies ≥ 6 par tumeur pour les gènes *ERBB2*, *MYC* et *ZNF703*.

d) Identification des mutations conductrices

Une combinaison de plusieurs outils est utilisée pour déduire les mutations conductrices de l'oncogenèse sur la base de méthodes statistiques par le biais de q-score : MutSigCV, OncodriveFML, MutPanning et FishHook. Nous avons choisi de combiner plusieurs outils pour augmenter la spécificité des résultats, et pour s'adapter aux petites cohortes sans perturber les analyses statistiques. Les gènes détectés comme driver plus d'une fois par l'un de ces outils avec un q-score < 0,05 et faisant partie des drivers connus dans l'oncogenèse mammaire(45), sont considérés comme un driver dans notre pipeline.

e) Signatures mutationnelles et détection de l'HRD

La plupart des études des signatures mutationnelles en analyse normal/tumeur utilisent des tissus tumoraux frais congelés exempts d'artéfact de séquençage FFPE. Des études similaires en analyse normal/tumeur avec des tumeurs FFPE sont rares, et il n'existe pas de pratique codifiée pour ce genre de protocole. C'est pourquoi nous avons choisi d'utiliser différents outils pour évaluer la signature mutationnelle à partir de données de WES : MutationalPattern, Maftools, et DeconstructSig pour le regroupement des signatures tumorales et la contribution de la signature mutationnelle dans chaque tumeur. Différentes approches sont utilisées : des extractions des signatures mutationnelles *de novo* par NMF et des analyses de comparaison aux signatures COSMIC à l'aide de la similarité cosinus.

Nous utilisons SigMA pour détecter les signatures mutationnelles associées à l'HRD et nous utilisons également ScarHRD pour détecter les cicatrices génomiques de l'HRD.

f) Algorithme de sélection de variants constitutionnels candidats associés à l'HRD :

Un algorithme de détection de variants constitutionnels a été développé en parallèle afin de découvrir de nouveaux gènes associés à l'HRD ou de reclassifier des VUS. Cet algorithme prend en compte les

annotations de ClinVar pour classer les variants constitutionnels déjà signalés, et lorsqu'aucune donnée ClinVar n'est disponible, nous utilisons les critères ACMG. Les variants classés comme probablement bénins ou bénins par ClinVar sont éliminés par filtrage.

Les variants faux-sens et de site d'épissage prédits comme étant délétères par au moins 4 de 8 outils de prédiction bio-informatique (SIFT, PolyPhen2, MutationTaster, FATHM, Provean, MetaSVM, MetaLR et CADD) sont pris en compte pour une analyse plus approfondie, ainsi que les variants tronquants (variants non-sens et avec décalage du cadre de lecture). À partir de cette liste de gènes candidats, nous recherchons des preuves de déséquilibre allélique à chaque locus de gènes dans la tumeur via les résultats de FACET sur l'état du locus (délétion/LOH par copie neutre) et via le compte de reads en utilisant des tests statistiques (test du Khi-deux) puis nous filtrons les variants dont la valeur p est >0,05.

B. Sélection des échantillons

1. En vue de la validation du pipeline sur une cohorte de patientes atteintes de cancer du sein.

Cette étude a été approuvée par un comité d'éthique de recherche local et des formulaires de consentement ont été recueillis. Nous avons identifié plus de 70 patientes candidates éligibles pour cette étude. Toutes présentaient un cancer du sein d'apparition précoce ou des antécédents familiaux importants de cancer du sein (score médian de Manchester : 24, intervalles 10-59). Aucune n'était porteuse d'un variant pathogène ou probablement pathogène dans les gènes *BRCA1/2* ou *PALB2* lors de précédents tests génétiques de routine. Nous avons examiné environ 130 blocs d'anatomo-pathologie afin de caractériser du matériel tumoral de cancer du sein disponible et adapté à l'analyse. Les échantillons de sang n'ont été prélevés que chez les patientes dont le matériel tumoral avait été récupéré et traité avec succès. Sur 59 tumeurs FFPE dont l'ADN a été extrait, 20 échantillons n'ont pu être séquencés en raison de la mauvaise qualité ou d'une quantité insuffisante d'ADN. Les WES ont été réalisés sur 39 paires normal/tumeur de femmes atteintes de cancer du sein diagnostiquées entre

1995 et 2017, mais nous en avons retiré une en raison d'un problème de contamination non résolu. L'intervalle entre le diagnostic et le séquençage est de quelques mois à 22 ans (médiane 3 ans).

L'ADN génomique a été extrait à partir des lymphocytes du sang périphérique des patientes et de la tumeur en utilisant des protocoles commerciaux publiés et des kits optimisés (kits Maxwell RSC Buffy Coat / Tissue DNA, Promega). La préparation de la bibliothèque, la capture de l'exome et le séquençage ont été menés sur une période de deux ans, au fur et à mesure de l'obtention du matériel, reflétant des essais répétés pour les échantillons d'ADN qui ont échoué. Un PoN composé de 47 échantillons de tissus normaux, tous séquencés selon le même protocole que celui appliqué aux tissus tumoraux, a été utilisé pour aider à éliminer les artefacts techniques récurrents. Les bibliothèques d'ADN génomique ont été générées à l'aide de l'une des méthodes suivantes : Nextera Rapid Capture Exome Target (Illumina), Sure Select Human All Exon V6, ou Sure Select Human All Exon V7 (Agilent). Les échantillons d'ADN ont ensuite été séquencés à l'aide des séquenceurs Illumina et des protocoles appropriés disponibles au moment du séquençage : Illumina HiSeq 2000 avec des reads de 100 pb en paires-end, Illumina HiSeq 2500 avec des reads de 125 pb en paires-end et Illumina HiSeq 4000 avec des reads de 100 pb en paires-end (Illumina).

2. En vue de l'application du pipeline en routine sur une cohorte de patientes atteintes de cancer de l'ovaire.

Afin d'évaluer notre outil de détection de l'HRD nous avons analysé 22 patientes atteintes de HGSC et naïves de chimiothérapie suivies à l'Institut Bergonié entre 2017 et 2020. 11 patientes ont précédemment été détectées pour être porteuses d'une mutation pathogène constitutionnelle des gènes *BRCA1* ou *BRCA2*, et ont constitué notre pool de cas positifs fortement susceptibles d'exprimer un phénotype tumoral HRD. 11 autres patientes non porteuses de mutation constitutionnelle des gènes *BRCA1* ou *BRCA2* ont constitué notre pool négatif. La plupart des outils de détection de l'HRD par approche NGS nécessitent une analyse à partir de paires de tissu sanguin et de tissus tumoraux

frais d'une même patiente et sont pour la plupart réalisés à partir de WES ou WGS, difficilement généralisables en routine.

Du fait de la difficulté d'obtenir du sang de patientes de manière rétrospective ou prospective, nous proposons ici de réaliser l'analyse sur des paires de tissus FFPE normaux (ganglions, tissus sains) et tumoraux. Cette technique ayant précédemment été démontrée comme similaire à une analyse normal/tumeur issu de sang et de tumeur(127).

Afin de pallier la perte de quantité de mutations secondaires à l'usage de petits panels de gènes, nous avons augmenté la sensibilité en utilisant le kit SureSelectXT Low Input (Agilent) afin d'obtenir une profondeur de 5000X. Pour nous rapprocher des conditions de routine nous avons réalisé le séquençage normal/tumeur sur notre panel de routine d'analyse constitutionnel de 65 gènes « panel NEIGE3 » et sur notre panel de routine d'analyse tumoral de 30 gènes somatiques du protocole GREAT, « panel GREAT+ ».

Ces deux paramètres que sont l'usage d'ADN issus de tissus FFPE et l'usage de panels restreint de 30 et 65 gènes, sont pour nous de véritables conditions de l'applicabilité clinique, et un véritable défi pour détecter des tumeurs HRD.

Un total de 22 ADN tumoraux fixés issus de pièces opératoires de HGSC et de 22 ADN normaux fixés issus de ganglions, de tissu sain adjacent, de myomètre ou de rate accessoire des patientes correspondantes, ont ensuite été quantifiés par Qubit (Illumina) en vue du séquençage. Deux échantillons présentaient une concentration en ADN plus faible que la concentration optimale nécessaire (< 200 ng/ml) et ont donc été moins dilués.

L'ADN génomique a été extrait du tissu sain des patientes sélectionnées et des tumeurs appareillées en utilisant des protocoles commerciaux publiés et des kits optimisés dans notre laboratoire. Les bibliothèques d'ADN génomique ont été générées à l'aide du kit SureSelectXT Low Input (Agilent). Les échantillons d'ADN ont ensuite été séquencés sur séquenceur Illumina NextSeq 500 (Illumina), et suivant les protocoles appropriés disponibles au moment du séquençage.

III. Résultats

A. Validation du pipeline

1. Démographie mutationnelle constitutionnelle

Grâce à ce pipeline, nous avons trouvé 1022 variants rares dans 279 gènes (moyenne par patientes de 26,2) dont 84,4% étaient des faux-sens et 7,53% des mutations tronquantes (4,11% des non-sens, et 3,42% des mutations par décalage du cadre de lecture). Après avoir appliqué notre algorithme de détection de variants constitutionnels, 168 variants ont été évalués. Trois variants ont été considérés comme pathogènes par ClinVar : *FANCA* c.2524delT ; p.Ser842LeufsTer47, *GJB2* c.35delG ; p.Gly12ValfsTer2 et *LIG4* c.2440C>T ; p.Arg814Ter. Nous avons trouvé 18 variants avec un *second hit* dans 16 gènes (*APC, ATM, BRDT, BRCA2, CABIN1, CHEK2, DNAH6, EIF2D, FLCN, LIG4, NCOR1, POLQ, RAD51D, REV1, TANGO2* et *XIRP2*)(128).

2. Démographie mutationnelle somatique

Nous avons trouvé 2297 mutations somatiques codantes dans les tumeurs, dont 82,5% étaient des mutations faux-sens, et 10% des mutations tronquantes (4% non-sens et 6% des mutations décalage du cadre de lecture). Des mutations conductrices dans les gènes associés aux cancers du sein ont été retrouvées dans 84% des tumeurs. Cependant, seules les mutations dans les gènes *TP53, GATA3, TBX3* et *CBFB* ont été considérées comme des mutations conductrices et présentes dans 21 échantillons (54%) par la combinaison des outils de détection des mutations conductrices(128). Conformément à la littérature, *TP53* était le gène le plus fréquemment muté et était presque toujours associé à une LOH.

3. Détection des tumeurs HRD et de variants candidats

Dans notre cohorte de 38 patientes atteintes de cancer du sein, nous avons identifié 7 tumeurs HRD avec SigMA, et 12 avec ScarHRD(128)(**Figure 28**). Les 7 cas présentant une signature HRD par l'analyse SigMA ont également été classés comme ayant des cicatrices génomiques HRD par ScarHRD. La seule tumeur HRD avec un variant candidat constitutionnel provenait d'une patiente porteuse d'un variant

faux-sens dans le gène *RAD51D* p.Ser207Leu (cas C10758). Grâce à l'analyse du nombre de copies, nous avons pu identifier la duplication de ce variant ainsi que la perte de l'allèle normal en *trans*, ce qui a entraîné une perte d'hétérozygotie par copie neutre au niveau du locus du gène comme deuxième évènement. Nous avons trouvé une délétion somatique sur un site donneur d'épissage dans le gène *RAD51C* associée à une duplication-LOH somatique dans une autre tumeur HRD (cas JGH1), mais sans autre événement associé. Aucune autre tumeur HRD n'a été retrouvée associée à des gènes candidats, que ce soit au niveau somatique ou constitutionnel. Ce travail de validation a fait l'objet d'une publication jointe au manuscrit en annexe (**Annexe 1**(128)).

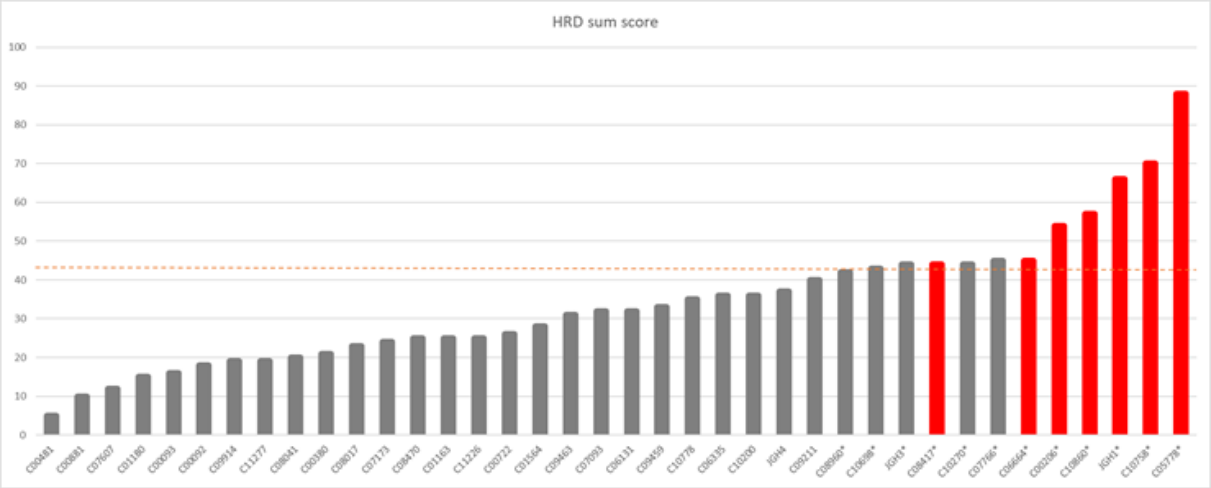
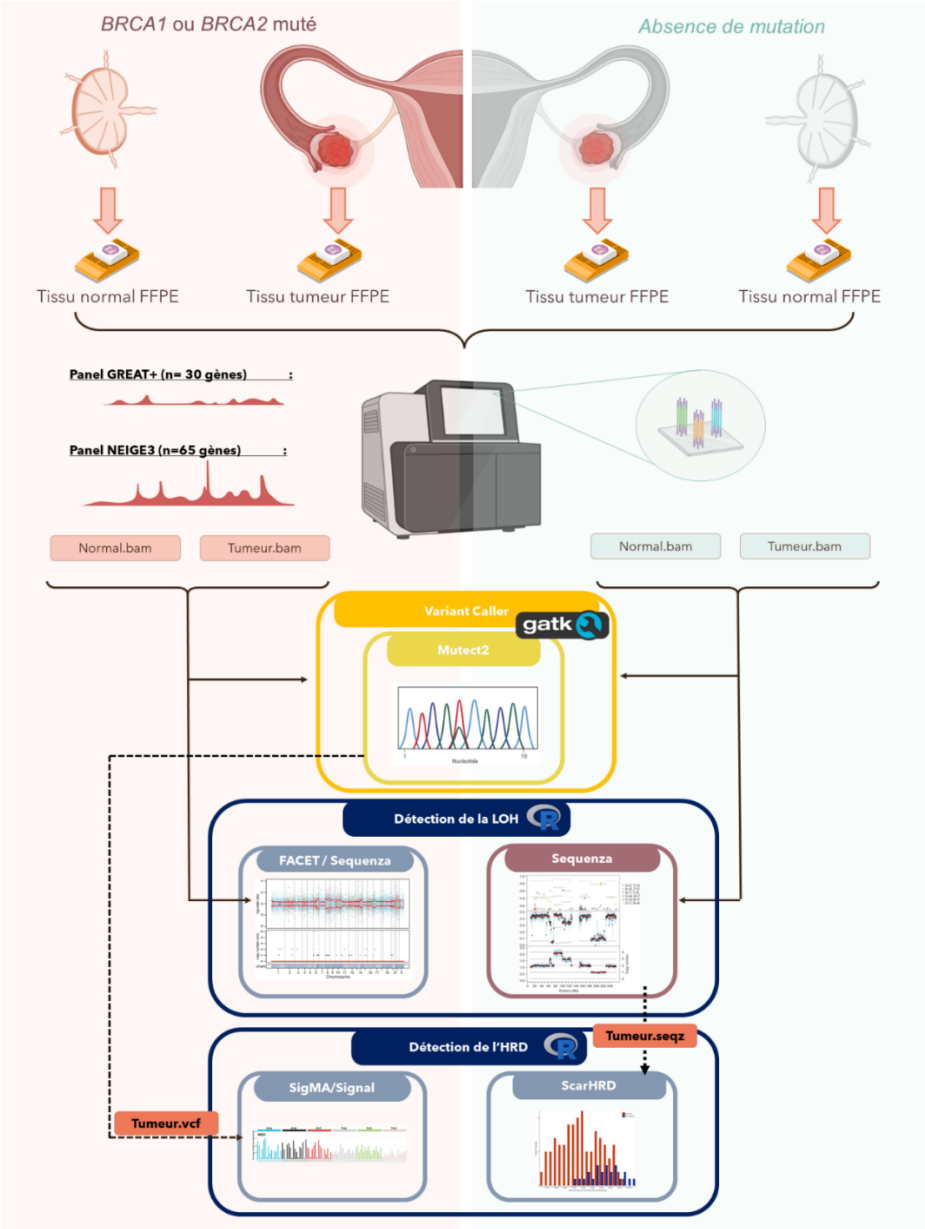


***Figure 28 :* Détection des tumeurs HRD en utilisant SigMA et ScarHRD dans la cohorte de cancer du sein.** *Un score « HRD sum Score » ≥ 42 définissait une tumeur HRD (délimité par la ligne en pointillé orange). Les tumeurs avec un score ≥ 42 sont marquées par un « * ». En rouge, sont marquées les tumeurs HRD selon SigMA.*

B. Structure du pipeline de routine

Pour une extrapolation de la détection d'une signature HRD en routine, nous avons légèrement modifié notre pipeline initial afin de l'adapter aux techniques utilisées dans notre laboratoire de diagnostic. Tous les outils décrit ci-dessous ont été rassemblés autour d'un nouveau pipeline bio-informatique développé pour l'occasion, sur notre serveur interne (**Figure 29).**



<u>Figure 29 :</u> Vignette du pipeline bio-informatique développé pour la détection de l'HRD en routine

Les variants somatiques ont été appelés individuellement en utilisant le mode normal/tumeur de Mutect2 pour les SNV et les indels. Les variants ont donc été éliminés par filtrage s'ils (i) présentaient un biais de brin, afin d'éliminer les artefacts causés par le processus de fixation au FFPE ; (ii) étaient d'origine constitutionnelle ; (iii) étaient dus à des artefacts techniques révélés par la comparaison avec le PoN, et (iv) avaient une profondeur < 10 reads. Le logiciel Alamut a été utilisé pour vérifier la présence des variants sur les fichiers Bam. Le mode *tumeur-only* de Mutect2, c'est-à-dire sans implémentation des BAM normaux, a été utilisé pour la tumeur T-POS6, afin d'évaluer si nous pouvions détecter les signatures SBS10a et SBS10b en s'affranchissant du contingent normal (voir partie III.C.4).

Pour l'exploration des CNA, nous avons utilisé à la fois FACETS et Sequenza. Pour le logiciel FACET, nous avons splitté la liste des SNP connues sur nos régions de captures afin de ne pas provoquer d'artéfact, sans prendre en compte les *off-targets.*

Nous avons utilisé le logiciel SigMA pour détecter les signatures mutationnelles liées à l'HRD, le logiciel ScarHRD afin de détecter les cicatrices génomiques de l'HRD et l'outil en ligne Signal pour détecter les autres signatures mutationnelles décrites dans la littérature.

C. Extrapolation et résultats du pipeline en routine

1. Démographie mutationnelle somatique

Un total de 22 paires normal/tumeur a été séquencées en utilisant les panels de capture GREAT+ et NEIGE3 à partir de tissu sain et de tumeurs. L'échantillon T-POS04 a été retiré de l'analyse car l'échantillon normal s'est avéré être muté (porteur du variant *TP53* retrouvé en tumoral), en faveur de la sélection de contingent tumoral à la place de tissu sain. L'échantillon T-NEG08 n'a pu être séquencé avec le panel de capture GREAT+. Un total de 20 paires a donc été pris en compte avec l'analyse GREAT+ et 21 paires avec NEIGE3. Avec le panel GREAT+ nous avons retrouvé 63 variants codants somatiques dont 57% de faux-sens, et 34% de mutations tronquantes, avec une moyenne de 3 variants par tumeur. Avec le panel NEIGE3, 82 variants codants somatiques ont été retrouvés, dont 61% étaient des faux-sens et 30% des mutations tronquantes, avec une moyenne de 4 variants par tumeur (**Annexe 2**). Plus de 90% des tumeurs étaient porteuses de mutation *TP53* en accord avec le phénotype HGSC des tumeurs de l'ovaire.

2. Rendement diagnostique de Mutect2 versus STARK

Cette analyse en normal/tumeur a permis de mettre en évidence que 98% des variants remontés dans nos analyses somatiques avec notre pipeline de routine diagnostic « STARK » à l'Institut Bergonié sont en réalité des variants constitutionnels, retrouvés dans la tumeur. Seulement 2% des variants sont de véritables variants somatiques, et ceci indépendamment du panel de capture utilisé.

La majorité des mutations somatiques identifiées par STARK a été retrouvée par Mutect2. L'ensemble des mutations retrouvées par Mutect2 et STARK a été contrôlé manuellement sur le logiciel Alamut. Hormis 2 variants d'*EPCAM* avec le panel GREAT+, Mutect2 n'a appelé aucun artéfact de séquençage. A l'inverse de STARK, ces variants ont été filtrés grâce à l'algorithme interne de Mutect2 en adéquation avec la bonne spécificité de Mutect2(105). En revanche 6 variants n'ont pas été appelés par Mutect2 alors que ces variants sont présents sur les BAMs et ne sont pas des artéfacts de séquençage (**Annexe 3**). Ceci illustre la moins bonne sensibilité de Mutect2 par rapport à STARK, qui en revanche appelle plus de variants, dont des variants absents des BAM, et a donc une moins bonne spécificité (**Tableau 4**). Combinés, ces 2 outils permettraient un appel des variants plus robuste et plus exhaustif. Ce procédé nécessiterait d'être comparé à un set de validation connu.

	Panel	Se	Sp
Mutect2	GREAT+	**0,9**	**0,78**
	NEIGE3	**0,88**	**1**
STARK	GREAT+	**1**	**0,33**
	NEIGE3	**0,96**	**0,33**

Tableau 4 : Comparaison de la sensibilité (Se) et spécificité (Sp) de Mutect2 et STARK selon le type de panel de capture utilisé dans notre cohorte de cancers de l'ovaire.

3. Détection de la signature HRD

Aucune tumeur n'a obtenu de score SigMA-HRD > 0,8 à la fois dans le contingent muté et non muté, et aucune différence n'a pu être observée entre le score SigMA-HRD et le statut mutationnel (**Figure 30**).

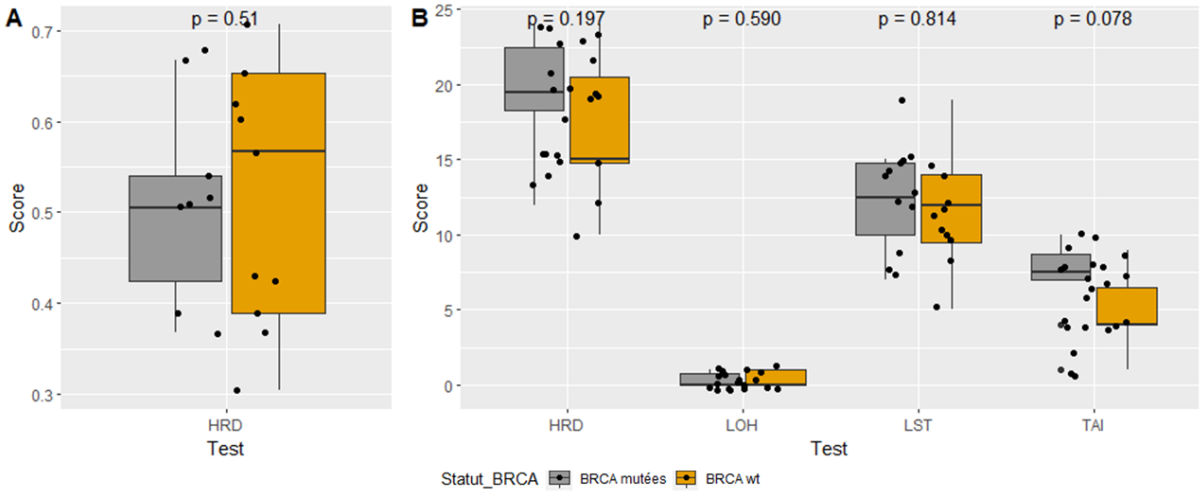


***Figure 30 :* Boxplots des scores HRD obtenus à l'aide de SigMA(A) et de ScarHRD(B).**

De plus, aucune tumeur n'a obtenu de score ≥ 42 avec ScarHRD, à la fois dans le contingent muté et non muté, et aucune différence n'a été observée dans les sous-scores TAI, LST et LOH, bien qu'une légère tendance puisse s'observer avec le sous-score TAI (**Figure 30**).

4. Détection d'une tumeur *POLE*-mutée

En revanche nous avons pu extraire les signatures SBS10a, SBS10b et SBS28 d'une tumeur présentant une forte charge mutationnelle (>100mut/Mb) (**Figure 31**).



***Figure 31 :* Profil mutationnel de la tumeur T-POS06 correspondant au phénotype mutationnel POLE.**

Ces signatures sont retrouvées dans les tumeurs mutées pour le gène *POLE*(131). Le gène *POLE* code la protéine ADN polymérase epsilon, qui participe à la correction des erreurs de réplication(132). Des mutations somatiques dans le domaine exonucléase de la *POLE* ont été identifiées dans environ 10% des cancers de l'endomètre, dans 1-2% des cancers colorectaux, et à de faibles fréquences dans d'autres type de cancer (cancer de l'estomac, glioblastome, cancer du sein). Les mutations somatiques récurrentes sont localisées dans un *hotspot* situé au niveau du domaine exonucléase codé par les exons 9 à 14, dont les plus fréquentes sont p.Pro286Arg, p.Ser297Phe et p.Val411Leu. Ces tumeurs sont caractérisées par un phénotype dit « *hypermutator* », c'est-à-dire une forte charge mutationnelle. Les tumeurs *POLE*-mutées présentent tout particulièrement une proportion élevée de substitution C>A et T>G, caractéristique de la signature SBS10a et SBS10b. La signature SBS28, d'étiologie inconnue, est fréquemment retrouvée comme signature accompagnatrice des signatures SBS10a et SBS10b. Les patientes atteintes d'un cancer de l'endomètre induit par une mutation délétère du gène *POLE* présentent un pronostic favorable, et peuvent bénéficier d'une immunothérapie. Cette tumeur a été initialement identifiée comme une Tumeur de Brenner où une recherche de mutation *BRCA* pour une éligibilité aux PARPi avaient retrouvé un nombre incohérent de variants non-sens et faux sens dans les gènes *BRCA1* et *BRCA2.* Un screening moléculaire tumoral avait été entrepris à l'époque, retrouvant la mutation du gène *POLE* p.Ser297Phe, et confirmé par notre nouveau pipeline. A la lumière de ces éléments, cette tumeur a été reclassée comme carcinome ovarien endométrioïde, sensible à l'immunothérapie.

Enfin, nous avons souhaité évaluer si nous pouvions nous affranchir de contingent normal en utilisant seulement la liste des SNP constitutionnels listée par gnomAD comme filtre et implémentée dans Mutect2. Nous avons ainsi comparé les mutations remontées par Mutect2 avec le mode normal/tumeur et *tumeur-only* pour la tumeur T-POS06, puis comparé le profil de signature mutationnelle.

Un total de 261 variants a été retrouvés pour 75 variants après filtrage grâce au mode normal/tumeur, alors que 319 variants ont été retrouvés pour 74 variants après filtrage avec le mode *tumeur-only*. Les signatures SBS10a et SBS10b ont été retrouvées dans les deux conditions, démontrant l'efficacité de Mutect2 même en l'absence de contingent normal (**Figure 32**).

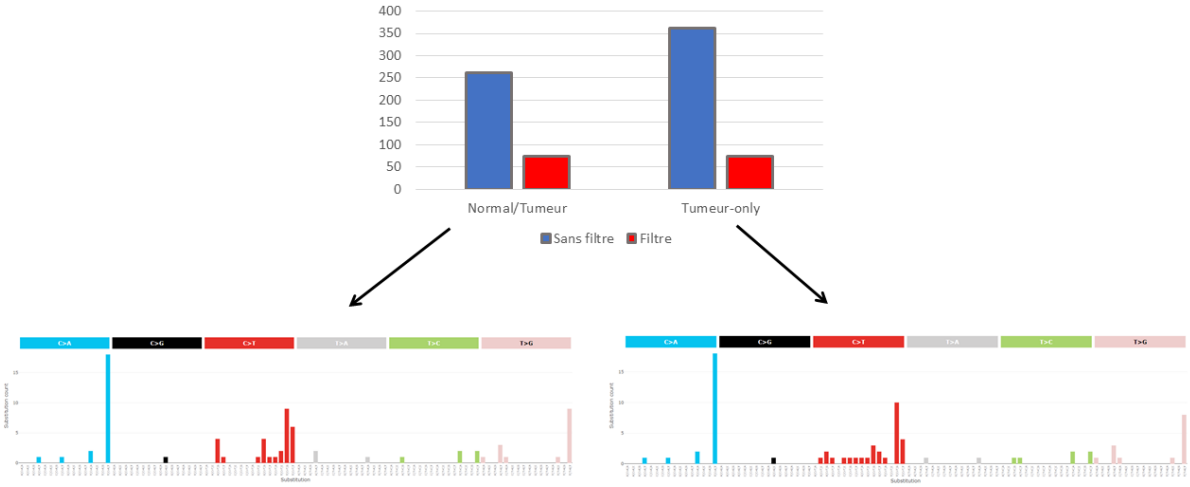


***Figure 32 :* Comparaison du nombre de variants remonté sans et après filtre selon utilisation du mode normal/tumeur ou tumeur-only.** *Les deux profils mutationnels sont quasi-identiques, retrouvant le profil des signatures SBS10a et SBS10b.*

IV. Discussion

Les résultats de ces deux projets démontrent qu'il est possible de détecter une signature HRD par analyse des signatures mutationnelles et des cicatrices génomiques issues d'échantillons FFPE grâce à une analyse normal/tumeur et au séquençage par WES. En revanche, aucune signature de l'HRD n'a pu être extraite avec les deux panels de routine. Les signatures mutationnelles associées à l'HRD n'ont pu être capturées car ces deux panels de gènes ne permettent pas de détecter assez d'événements mutationnels nécessaires à la détection de ces signatures. Bien que le nombre de variants soit un facteur quantitatif nécessaire à la détection d'un contexte mutationnel représentatif, il ne semble pas être corrélé à une signature HRD qui n'est pas de type *hypermutator* (**Figure 33**). C'est bien le type de SNV pris dans leur contexte tri-nucléotidique plutôt que le nombre qui permet d'extraire une signature SBS3.

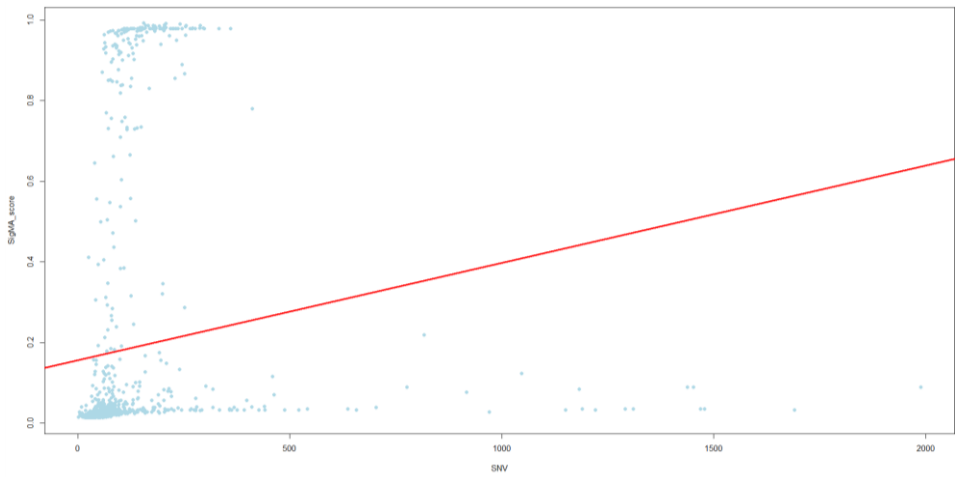


***Figure 33 :* Scatter-plot entre le nombre de variants (SNV) et le score SigMA-HRD (SigMA_score) issus de 1022 cancers du sein des données du TGCA**. *Le score de corrélation de Pearson (r^2 = 0,13, p <0.0001) est en faveur d'une très faible corrélation entre la charge mutationnelle d'une tumeur et les signatures mutationnelles associées à l'HRD.*

De plus, les deux outils de détection des CNA et LOH ont été conçus pour des analyses par WGS et/ou WES. Ils peuvent ne pas être aussi performant avec une analyse en panel de gènes. Enfin, le score de détection de l'HRD et les sous-scores ne sont peut-être pas adaptés pour ce genre d'analyse. Aucune tumeur de la cohorte, indépendamment du statut BRCA, n'a de score LOH >1, reflétant un défaut de détection. Ce score est établi sur des pertes de plus de 15Mb sur un chromosome. Or notre pipeline

ne capturant que des zones restreintes de quelques kilobases, il ne permet pas l'évaluation de tels réarrangements. Il n'est donc pas possible d'utiliser un outil de détection de l'HRD, avec les techniques actuelles.

La majorité des études évaluant les signatures HRD sont basées sur des analyses normal/tumeur issues de tissus frais congelés(53,54,92,93). De telles analyses sont difficiles à mettre en œuvre en routine diagnostique bien qu'ayant de très bons résultats. Cependant, en appliquant une analyse combinée constitutionnelle et somatique, associée à la détection des CNA et de LOH, ainsi que l'analyse d'une signature HRD, nous avons pu approfondir l'analyse de VUS et des variants pathogènes ou probablement pathogènes pour les agrégations familiales de cancers du sein et ou de l'ovaire, et développer un pipeline bio-informatique permettant la détection des tumeurs HRD FFPE. Cette approche a déjà été testée sur une cohorte de cancers du sein à haut risque par séquençage du génome entier ou WES, mais elle ne portait que sur les gènes de prédisposition aux cancers du sein(102,133). De plus, ici, nous avons pu analyser des échantillons FFPE provenant d'archives hospitalières, remontant jusqu'à plus de 20 ans. De plus, il a été démontré récemment qu'en utilisant un filtre adapté via MuTect2 comme nous l'avons fait, même des échantillons FFPE issus de ganglions ou de tissus sains adjacents pouvaient être utilisés comme une ressource constitutionnelle pour l'analyse en normal/tumeurs, ce que nous confirmons(127). Enfin, l'utilisation d'un panel de gènes plus large ou d'une banque de SNP, mieux répartis tout le long du génome, ainsi que d'un calcul du nombre de copies, de reads et de la LOH exclusivement centré sur les gènes du panel pourraient être une bonne alternative par rapport aux WES pour détecter les signatures associées à l'HRD.

Détecter une signature HRD présente un intérêt diagnostique et économique car la prescription d'inhibiteur de PARP en traitement de maintenance dans le cancer de l'ovaire séreux de haut grade n'est possible que si les patientes sont atteintes de tumeurs porteuses de mutations perte de fonction dans les gènes *BRCA1/2* ou d'une signature HRD, actuellement détectée par le test breveté MyChoice® de Myriad Genetics. Ce test étant basé sur la détection des cicatrices génomiques de l'HRD, plusieurs

laboratoires français du groupe Unicancer se sont lancés dans le développement en interne de tests HRD basés sur des analyses des CNA et de l'instabilité génétique par CGH-array, SNP-array ou panel de gènes. Aucune étude clinique n'a pour l'instant validé l'utilisation des signatures mutationnelles, ni d'autres tests fonctionnels dans l'éligibilité des tumeurs aux inhibiteurs de PARP(134).

La détection des tumeurs présentant une signature HRD est un défi auquel les laboratoires d'oncogénétique et de biologie des tumeurs vont devoir répondre dans les années à venir. Environ 6% des cancers tous confondus présenteraient une signature HRD, et potentiellement sensibles aux inhibiteurs de PARP. Des essais cliniques sont donc nécessaires pour évaluer l'efficacité de ces molécules dans les différents organes.

L'éligibilité à un traitement par inhibiteur de PARP est conditionnée par le dogme qu'une tumeur HRD y serait sensible. Ce constat a amené l'étude Olympia à élargir l'indication de l'olaparib en adjuvant et néo-adjuvant dans le cancer du sein localisé Her2- sous condition de la positivité du test HRD breveté MyChoice® comme seul test diagnostique(100). Or la détection des cicatrices génomiques et de ce test, présentent toutefois quelques limites dans leur attribution d'une signature HRD et donc de la prédiction qu'une tumeur serait sensible aux PARPi. La première, est que ces tests ont une moins bonne spécificité pour prévoir la sensibilité aux sels de platine lorsque les patientes ont été traitées antérieurement par anthracyclines. Une exposition antérieure aux anthracyclines dans ce contexte, peut contribuer à l'inexactitude du test suite à l'émergence de voies de résistance qui n'éliminerait pas les " cicatrices génomiques " existantes de l'HRD(74). La seconde, sont les possibles réversions ou secondes mutations dans les gènes *BRCA1/2*, *PALB2*, et *RAD51C/D* restaurant la fonctionnalité de la voie HR sans corriger les conséquences antérieures de l'HRD sur le génome. Ceci est particulièrement pertinent dans le cas du cancer de l'ovaire, où environ la moitié des tumeurs mutées *BRCA* résistantes aux sels de platine finissent par restaurer la fonction des protéines BRCA(135). Les autres mécanismes de résistance comprennent la stabilisation des mutants du domaine BRCT de BRCA1 par HSP90, l'épissage alternatif, ou l'initiation alternative de la traduction de *BRCA1*. La troisième est la résistance

acquise ou intrinsèque fréquemment médiée par des transporteurs membranaires, comme cela a été décrit dans la résistance à l'olaparib dans des modèles de souris due à l'augmentation de la pompe d'efflux P-glycoprotéine. Enfin la dernière limite réside dans la sensibilité du test lui-même. Dans les essais cliniques PAOLA, et PRIMA, l'olaparib et le niraparib ont amélioré la survie sans progression (PFS), même dans les cohortes classées comme HRP(98,99). Or si l'on considère que les tumeurs sensibles aux PARPi ne peuvent s'expliquer que par la présence d'une signature HRD, ces dernières sont donc manquées par le test MyChoice®. Ce manque de sensibilité peut nous amener à nous poser la question « l'HRD est-il le test le mieux indiqué pour évaluer la réponse aux inhibiteurs de PARP ?».

Actuellement au laboratoire, seule la recherche constitutionnelle et/ou somatique de variant perte de fonction dans les gènes *BRCA1/2* permet de classer les tumeurs dans la catégorie HRD et donc sensibles aux inhibiteurs de PARP. Cette recherche pouvant être orientée par la positivité du test MyChoice®. Cependant, cette approche présente plusieurs inconvénients. Tout d'abord elle ne permet pas de détecter les tumeurs BRCAness. Ensuite la mise en évidence de VUS dans la tumeur nécessite d'y ajouter des études fonctionnelles pour trancher sur leur pathogénicité. Enfin, les mutations somatiques de réversion peuvent rendre une tumeur fonctionnellement compétente en matière de recombinaison homologue, même en présence de mutations constitutionnelles dans les gènes liés à la réparation par recombinaison homologue(136).

Parmi les autres techniques disponibles et largement utilisées dans les laboratoires, les analyses par CGH-array ont déjà montré une sensibilité et une spécificité cohérentes dans la détection des tumeurs HRD. Des profils tumoraux de tumeurs mutées *BRCA1* ou *BRCA2* ont été extraits et combinés pour établir un score BRCA-likeCGH, validé sur des tumeurs FFPE. A noter que deux tiers seulement des tumeurs de type BRCA1 étaient porteuses de mutation, suggérant que cette signature peut capturer un spectre plus large de tumeurs HRD. Ce score BRCA-likeCGH a été évalué rétrospectivement dans le cadre d'un essai clinique sur le cancer du sein comparant chimiothérapie seule à une chimiothérapie suivie d'une dose élevée de platine et une autogreffe de cellules souches. L'étude a montré que les

patients BRCA-likeCGH positif ont bénéficié d'un avantage statistiquement significatif en termes de survie par traitement à base de platine. Cependant, les tests par CGH-array n'ont pas été évalués dans le contexte des inhibiteurs de PARP(137,138). A noter l'existence de tests fonctionnels évaluant la formation des foci RAD51 après irradiation comme excellent marqueur de la voie HR, mais nécessitant du tissu frais, et une expertise, actuellement réservés qu'à quelques centres.

Pour l'heure, contrairement aux altérations génomiques qui sont le reflet d'événements passés, l'ARN et les protéines varient dynamiquement en quantité et en localisation au cours des processus cellulaires. Ils sont donc supposés fournir en instantané l'état de la voie HR dans une tumeur. Ainsi, pour la plupart des limites mentionnées ci-dessus, les tests qui rapportent un statut "en temps réel" de la voie HR (Real-time index of HRD) dans une tumeur, seraient d'intérêt et de valeur thérapeutique. Dans le cancer du sein et de l'ovaire, les tests d'expression génique sur puce à ADN (gene microarray-based assay) ont permis d'extraire des profils tumoraux liés à un défaut dans la voie HR dans les cancers génétiquement favorisés (gBRCA), et dans les cancers sporadiques (BRCAness). Ces profils BRCA ont démontré une bonne efficacité pour prédire la sensibilité aux sels de platine mais aussi l'apparition de résistance au cours du traitement, se reflétant par un changement de profil, soulignant la nature dynamique de ces tests(139,140). Aucun de ces tests n'a pour l'instant été utilisé dans des essais cliniques de PARPi.

Pour conclure, la généralisation des tests génétiques chez les personnes atteintes ou à risque de cancer du sein/des ovaires et d'autres cancers ont conduit à la découverte de VUS, dont beaucoup sont "privés" avec une fréquence en population assez faible. Les familles étant de plus en plus petites, les analyses de ségrégation sont difficiles à entreprendre, les analyses fonctionnelles ne sont pas toujours possibles et les prédictions *in silico* ne suffisent pas à elles seules à la prise de décision clinique. L'apport du séquençage tumoral et notamment de l'analyse en normal/tumeur permet d'ajouter des données phénotypiques (exemple : signature HRD, MMR, POLE) et oncogéniques (2^{nd} hit, LOH) pouvant apporter un poids supplémentaire dans la classification et l'interprétation des variants constitutionnels

et tumoraux (critère ACMG : PP4). Ces critères ne sont actuellement pas reconnus dans l'interprétation des variants du syndrome HBOC à l'inverse des variants des gènes MMR dans le syndrome de Lynch, et pourraient à l'avenir permettre de classer un certain nombre de VUS.

Afin d'améliorer sa spécificité et de trancher sur la pathogénicité d'un VUS, la recherche de variant constitutionnel dans les gènes *BRCA* devrait s'accompagner d'une analyse tumorale de ces gènes en parallèle, pour rechercher une rétention de variant somatique, témoin indirect d'un mécanisme de LOH sous-jacent, voire d'une analyse pan-génomique à la recherche directe de CNA et LOH. L'absence de LOH spécifique au locus BRCA, par exemple, est associée à une moins bonne réponse aux sels de platine dans les tumeurs mutantes gBRCA(141).

V. Conclusion

Bien qu'il soit possible de détecter une signature HRD au travers d'une analyse par WES, cette méthode n'est pas applicable en routine dans les laboratoires de biologie moléculaire utilisant les panels actuels recommandés pour le diagnostic.

En revanche, au travers de la détection de cette tumeur présentant une mutation du gène *POLE* p.Ser297Phe – SBS10a/b, ainsi que de la bonne corrélation des résultats somatiques identifiés précédemment avec notre pipeline de référence STARK, ce pipeline adapté à la routine :

- Est efficace pour détecter des signatures mutationnelles lorsqu'un nombre conséquent d'événement mutationnels sont détectés (>30),
- Ne remonte que très peu d'artéfacts et conserve une très bonne spécificité malgré l'utilisation exclusive de tissus FFPE,
- Fonctionne en utilisant aussi bien du tissu sain FFPE que des échantillons sanguins comme contingent normal,
- Pourrait être utilisé sans contingent normal pour détecter des signatures mutationnelles associées à de fortes charges mutationnelles, bien qu'un nouveau set de validation soit nécessaire pour s'assurer d'une bonne corrélation.

VI. Références

1. Kurian AW, Hughes E, Handorf EA, Gutin A, Allen B, Hartman A-R, et al. Breast and Ovarian Cancer Penetrance Estimates Derived From Germline Multiple-Gene Sequencing Results in Women. JCO Precis Oncol. 2017 Jun 27;(1):1–12.

2. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971 Apr;68(4):820–3.

3. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer. 2016 Feb;16(2):110–20.

4. Toland AE, Andreassen PR. DNA repair-related functional assays for the classification of BRCA1 and BRCA2 variants: a critical review and needs assessment. J Med Genet. 2017;54(11):721–31.

5. Scully R, Livingston DM. In search of the tumour-suppressor functions of BRCA1 and BRCA2. Nature. 2000 Nov 23;408(6811):429–32.

6. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017 Oct;49(10):1476–86.

7. Friedman LS, Ostermeyer EA, Szabo CI, Dowd P, Lynch ED, Rowell SE, et al. Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. Nat Genet. 1994 Dec;8(4):399–404.

8. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994 Oct 7;266(5182):66–71.

9. Cleton-Jansen AM, Collins N, Lakhani SR, Weissenbach J, Devilee P, Cornelisse CJ, et al. Loss of heterozygosity in sporadic breast tumours at the BRCA2 locus on chromosome 13q12-q13. Br J Cancer. 1995 Nov;72(5):1241–4.

10. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995 Dec 21;378(6559):789–92.

11. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer. 2011 Dec 23;12(1):68–78.

12. Sobhian B, Shao G, Lilli DR, Culhane AC, Moreau LA, Xia B, et al. RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. Science. 2007 May 25;316(5828):1198–202.

13. Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell. 2001 Apr 6;105(1):149–60.

14. Yu X, Wu LC, Bowcock AM, Aronheim A, Baer R. The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. J Biol Chem. 1998 Sep 25;273(39):25388–92.

15. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet. 2007 Feb;39(2):159–61.

16. Ahlskog JK, Larsen BD, Achanta K, Sørensen CS. ATM/ATR-mediated phosphorylation of PALB2 promotes RAD51 function. EMBO Rep. 2016 May;17(5):671–81.

17. Dosanjh MK, Collins DW, Fan W, Lennon GG, Albala JS, Shen Z, et al. Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. Nucleic Acids Res. 1998 Mar 1;26(5):1179–84.

18. Pittman DL, Weinberg LR, Schimenti JC. Identification, characterization, and genetic mapping of Rad51d, a new mouse and human RAD51/RecA-related gene. Genomics. 1998 Apr 1;49(1):103–11.

19. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol. 2015 Apr 1;7(4):a016600.

20. Kuznetsov SG, Haines DC, Martin BK, Sharan SK. Loss of Rad51c leads to embryonic lethality and modulation of Trp53-dependent tumorigenesis in mice. Cancer Res. 2009 Feb 1;69(3):863–72.

21. Pittman DL, Schimenti JC. Midgestation lethality in mice deficient for the RecA-related gene, Rad51d/Rad51l3. Genes N Y N 2000. 2000 Mar;26(3):167–73.

22. Savitsky K, Sfez S, Tagle DA, Ziv Y, Sartiel A, Collins FS, et al. The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. Hum Mol Genet. 1995 Nov;4(11):2025–32.

23. Cortez D, Wang Y, Qin J, Elledge SJ. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. Science. 1999 Nov 5;286(5442):1162–6.

24. Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, et al. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet. 1996 Dec;14(4):430–40.

25. Zhao W, Steinfeld JB, Liang F, Chen X, Maranon DG, Jian Ma C, et al. BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. Nature. 2017 Oct 19;550(7676):360–5.

26. Blasina A, de Weyer IV, Laus MC, Luyten WH, Parker AE, McGowan CH. A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. Curr Biol CB. 1999 Jan 14;9(1):1–10.

27. Lee JS, Collins KM, Brown AL, Lee CH, Chung JH. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. Nature. 2000 Mar 9;404(6774):201–4.

28. Falck J, Mailand N, Syljuåsen RG, Bartek J, Lukas J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. Nature. 2001 Apr 12;410(6830):842–7.

29. Huen MSY, Sy SMH, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. Nat Rev Mol Cell Biol. 2010 Feb;11(2):138–48.

30. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. Nat Rev Cancer. 2012 Dec;12(12):801–17.

31. Lavin MF, Kozlov S, Gatei M, Kijas AW. ATM-Dependent Phosphorylation of All Three Members of the MRN Complex: From Sensor to Adaptor. Biomolecules. 2015 Oct 23;5(4):2877–902.

32. Zhong Q, Chen CF, Li S, Chen Y, Wang CC, Xiao J, et al. Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. Science. 1999 Jul 30;285(5428):747–50.

33. San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Annu Rev Biochem. 2008;77:229–57.

34. Suwaki N, Klare K, Tarsounas M. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. Semin Cell Dev Biol. 2011 Oct;22(8):898–905.

35. Liu J, Doty T, Gibson B, Heyer W-D. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nat Struct Mol Biol. 2010 Oct;17(10):1260–2.

36. Jensen RB, Carreira A, Kowalczykowski SC. Purified human BRCA2 stimulates RAD51-mediated recombination. Nature. 2010 Oct 7;467(7316):678–83.

37. Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, et al. PALB2 links BRCA1 and BRCA2 in the DNA-damage response. Curr Biol CB. 2009 Mar 24;19(6):524–9.

38. Chun J, Buechelmaier ES, Powell SN. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. Mol Cell Biol. 2013 Jan;33(2):387–95.

39. Buisson R, Dion-Côté A-M, Coulombe Y, Launay H, Cai H, Stasiak AZ, et al. Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nat Struct Mol Biol. 2010 Oct;17(10):1247–54.

40. Scully R, Livingston DM. In search of the tumour-suppressor functions of BRCA1 and BRCA2. Nature. 2000 Nov 23;408(6811):429–32.

41. Li A, Geyer FC, Blecua P, Lee JY, Selenica P, Brown DN, et al. Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer. 2019;5:23.

42. Weigelt B, Bi R, Kumar R, Blecua P, Mandelker DL, Geyer FC, et al. The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J Natl Cancer Inst. 2018 01;110(9):1030–4.

43. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61–70.

44. Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, et al. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2008 Sep 10;26(26):4282–8.

45. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016 02;534(7605):47–54.

46. Kurman RJ, Shih I-M. The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded. Am J Pathol. 2016 Apr;186(4):733–47.

47. Bowtell DD, Böhm S, Ahmed AA, Aspuria P-J, Bast RC, Beral V, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer. 2015 Nov;15(11):668–79.

48. Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. Nat Rev Dis Primer. 2016 Aug 25;2:16061.

49. Shih I-M, Wang Y, Wang T-L. The Origin of Ovarian Cancer Species and Precancerous Landscape. Am J Pathol. 2021 Jan;191(1):26–39.

50. McCluggage WG. Ovarian neoplasms composed of small round cells: a review. Adv Anat Pathol. 2004 Nov;11(6):288–96.

51. Jelinic P, Mueller JJ, Olvera N, Dao F, Scott SN, Shah R, et al. Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. Nat Genet. 2014 May;46(5):424–6.

52. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. Nat Rev Cancer. 2004 Oct;4(10):814–9.

53. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017 Apr;23(4):517–25.

54. Nguyen L, W. M. Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun. 2020 Dec;11(1):5584.

55. Glodzik D, Bosch A, Hartman J, Aine M, Vallon-Christersson J, Reuterswärd C, et al. Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers. Nat Commun. 2020 Dec;11(1):3747.

56. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. Cancer Discov. 2015 Nov;5(11):1137–54.

57. Kast K, Rhiem K, Wappenschmidt B, Hahnen E, Hauke J, Bluemcke B, et al. Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. J Med Genet. 2016 Jul;53(7):465–71.

58. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. JAMA. 2017 Jun 20;317(23):2402.

59. Bolton KL, Chenevix-Trench G, Goh C, Sadetzki S, Ramus SJ, Karlan BY, et al. Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. JAMA. 2012 Jan 25;307(4):382–90.

60. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. J Clin Oncol Off J Am Soc Clin Oncol. 2020 Mar 1;38(7):674–85.

61. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. N Engl J Med. 2021 Feb 4;384(5):440–51.

62. Breast Cancer Association Consortium, Dorling L, Carvalho S, Allen J, González-Neira A, Luccarini C, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med. 2021 Feb 4;384(5):428–39.

63. Lhotova K, Stolarova L, Zemankova P, Vocka M, Janatova M, Borecka M, et al. Multigene Panel Germline Testing of 1333 Czech Patients with Ovarian Cancer. Cancers. 2020 Apr 13;12(4).

64. Song H, Dicks E, Ramus SJ, Tyrer JP, Intermaggio MP, Hayward J, et al. Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. J Clin Oncol Off J Am Soc Clin Oncol. 2015 Sep 10;33(26):2901–7.

65. Weber-Lassalle N, Hauke J, Ramser J, Richters L, Groß E, Blümcke B, et al. BRIP1 loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. Breast Cancer Res BCR. 2018 Jan 24;20(1):7.

66. Yang X, Song H, Leslie G, Engel C, Hahnen E, Auber B, et al. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. J Natl Cancer Inst. 2020 Feb 28;

67. Eisinger F, Bressac B, Castaigne D, Cottu P-H, Lansac J, Lefranc J-P, et al. [Identification and management of hereditary predisposition to cancer of the breast and the ovary (update 2004)]. Bull Cancer (Paris). 2004 Mar;91(3):219–37.

68. Moretta J, Berthet P, Bonadona V, Caron O, Cohen-Haguenauer O, Colas C, et al. [The French Genetic and Cancer Consortium guidelines for multigene panel analysis in hereditary breast and ovarian cancer predisposition]. Bull Cancer (Paris). 2018 Oct;105(10):907–17.

69. Brody LC. Treating cancer by targeting a weakness. N Engl J Med. 2005 Sep 1;353(9):949–50.

70. Byrski T, Huzarski T, Dent R, Marczyk E, Jasiowka M, Gronwald J, et al. Pathologic complete response to neoadjuvant cisplatin in BRCA1-positive breast cancer patients. Breast Cancer Res Treat. 2014 Sep;147(2):401–5.

71. Byrski T, Dent R, Blecharz P, Foszczynska-Kloda M, Gronwald J, Huzarski T, et al. Results of a phase II open-label, non-randomized trial of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer. Breast Cancer Res BCR. 2012 Jul 20;14(4):R110.

72. Byrski T, Huzarski T, Dent R, Gronwald J, Zuziak D, Cybulski C, et al. Response to neoadjuvant therapy with cisplatin in BRCA1-positive breast cancer patients. Breast Cancer Res Treat. 2009 May;115(2):359–63.

73. Silver DP, Richardson AL, Eklund AC, Wang ZC, Szallasi Z, Li Q, et al. Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2010 Mar 1;28(7):1145–53.

74. Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P, et al. Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. Nat Med. 2018 May;24(5):628–37.

75. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature. 2012 Jan 18;481(7381):287–94.

76. Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005 Apr 14;434(7035):917–21.

77. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005 Apr 14;434(7035):913–7.

78. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. Nat Rev Mol Cell Biol. 2017 Oct;18(10):610–21.

79. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005 Apr 14;434(7035):913–7.

80. Castéra L, Krieger S, Rousselin A, Legros A, Baumann J-J, Bruet O, 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 EJHG. 2014 Nov;22(11):1305–13.

81. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005 Sep 9;309(5741):1728–32.

82. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

83. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci. 2010 Jul 13;107(28):12629–33.

84. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 2007 Feb;39(2):165–7.

85. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010 May;42(5):410–4.

86. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008;29(11):1282–91.

87. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 Off J Am Coll Med Genet. 2015 May;17(5):405–24.

88. Calò V, Bruno L, La Paglia L, Perez M, Margarese N, Di Gaudio F, et al. The Clinical Significance of Unknown Sequence Variants in BRCA Genes. Cancers. 2010 Sep 10;2(3):1644–60.

89. Federici G, Soddu S. Variants of uncertain significance in the era of high-throughput genome sequencing: a lesson from breast and ovary cancers. J Exp Clin Cancer Res. 2020 Mar 4;39(1):46.

90. Sztupinszki Z, Diossy M, Börcsök J, Prosz A, Cornelius N, Kjeldsen MK, et al. Comparative Assessment of Diagnostic Homologous Recombination Deficiency associated mutational signatures in ovarian cancer. Clin Cancer Res [Internet]. 2021 Jan 1 [cited 2021 Aug 13];

Available from: https://clincancerres.aacrjournals.org/content/early/2021/08/10/1078-0432.CCR-21-0981

91. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013 Jan 31;3(1):246–59.

92. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug 22;500(7463):415–21.

93. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012 May 25;149(5):979–93.

94. Birkbak NJ, Wang ZC, Kim J-Y, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer Discov. 2012 Apr;2(4):366–75.

95. Popova T, Manie E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with BRCA1/2 Inactivation. Cancer Res. 2012 Nov 1;72(21):5454–62.

96. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer. 2012 Nov 6;107(10):1776–82.

97. Moore K, Colombo N, Scambia G, Kim B-G, Oaknin A, Friedlander M, et al. Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N Engl J Med. 2018 Dec 27;379(26):2495–505.

98. Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, et al. Olaparib plus Bevacizumab as First-Line Maintenance in Ovarian Cancer. N Engl J Med. 2019 Dec 19;381(25):2416–28.

99. González-Martín A, Pothuri B, Vergote I, DePont Christensen R, Graybill W, Mirza MR, et al. Niraparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. N Engl J Med. 2019 Dec 19;381(25):2391–402.

100. Tutt ANJ, Garber JE, Kaufman B, Viale G, Fumagalli D, Rastogi P, et al. Adjuvant Olaparib for Patients with BRCA1- or BRCA2-Mutated Breast Cancer. N Engl J Med. 2021 Jun 24;384(25):2394–405.

101. Sulkowski PL, Oeck S, Dow J, Economos NG, Mirfakhraie L, Liu Y, et al. Oncometabolites suppress DNA repair by disrupting local chromatin signalling. Nature. 2020 Jun;582(7813):586–91.

102. Van Marcke C, Helaers R, De Leener A, Merhi A, Schoonjans CA, Ambroise J, et al. Tumor sequencing is useful to refine the analysis of germline variants in unexplained high-risk breast cancer families. Breast Cancer Res. 2020 Dec;22(1):36.

103. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013 Mar;31(3):213–9.

104. Krøigård AB, Thomassen M, Lænkholm A-V, Kruse TA, Larsen MJ. Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. PloS One. 2016;11(3):e0151664.

105. Cai L, Yuan W, Zhang Z, He L, Chou K-C. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. Sci Rep. 2016 Nov 22;6:36540.

106. de Schaetzen van Brienen L, Larmuseau M, Van der Eecken K, De Ryck F, Robbe P, Schuh A, et al. Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. BMC Med Genomics. 2020 Jul 6;13(1):94.

107. Robbe P, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med Off J Am Coll Med Genet. 2018 Oct;20(10):1196–205.

108. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013 Jul;499(7457):214–8.

109. Dietlein F, Weghorn D, Taylor-Weiner A, Richters A, Reardon B, Liu D, et al. Identification of cancer driver genes based on nucleotide context. Nat Genet. 2020 Feb;52(2):208–18.

110. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol. 2016 Dec;17(1):128.

111. Imielinski M, Guo G, Meyerson M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell. 2017 Jan;168(3):460-472.e14.

112. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011 Apr;12(4):R41.

113. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res. 2016 19;44(16):e131.

114. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol Off J Eur Soc Med Oncol. 2015 Jan;26(1):64–70.

115. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018 Apr 25;10(1):33.

116. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016 Feb 22;17:31.

117. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018 Nov;28(11):1747–56.

118. Cyriac Kandoth. Vcf2Maf V1.5 [Internet]. Zenodo; 2015 [cited 2021 Jun 20]. Available from: https://zenodo.org/record/593251

119. Gulhan DC, Lee JJ-K, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet. 2019;51(5):912–9.

120. Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. NPJ Breast Cancer. 2018;4:16.

121. Timms KM, Abkevich V, Hughes E, Neff C, Reid J, Morris B, et al. Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. Breast Cancer Res BCR. 2014 Dec 5;16(6):475.

122. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2016 Aug 1;22(15):3764–73.

123. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010 Jan 14;463(7278):191–6.

124. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009 Oct 8;461(7265):809–13.

125. Wong SQ, Li J, Tan AY-C, Vedururu R, Pang J-MB, Do H, et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics. 2014 May 13;7:23.

126. Quach N, Goodman MF, Shibata D. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. BMC Clin Pathol. 2004 Feb 12;4(1):1.

127. Wilkins A, Chauhan R, Rust A, Pearson A, Daley F, Manodoro F, et al. FFPE breast tumour blocks provide reliable sources of both germline and malignant DNA for investigation of genetic determinants of individual tumour responses to treatment. Breast Cancer Res Treat. 2018 Aug;170(3):573–81.

128. Matis TS, Zayed N, Labraki B, de Ladurantaye M, Matis TA, Camacho Valenzuela J, et al. Current gene panels account for nearly all homologous recombination repair-associated multiple-case breast cancer families. Npj Breast Cancer. 2021 Aug 25;7(1):109.

129. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet. 2012 Jul;49(7):433–6.

130. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Dec;17(1):122.

131. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020 Feb;578(7793):94–101.

132. Davila JI, Chanana P, Sarangi V, Fogarty ZC, Weroha SJ, Guo R, et al. Frequent POLE-driven hypermutation in ovarian endometrioid cancer revealed by mutational signatures in RNA sequencing. BMC Med Genomics. 2021 Jun 22;14(1):165.

133. Nones K, Johnson J, Newell F, Patch AM, Thorne H, Kazakoff SH, et al. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. Ann Oncol Off J Eur Soc Med Oncol. 2019 01;30(7):1071–9.

134. Hoppe MM, Sundar R, Tan DSP, Jeyasekharan AD. Biomarkers for Homologous Recombination Deficiency in Cancer. J Natl Cancer Inst. 2018 Jul 1;110(7):704–13.

135. Norquist B, Wurz KA, Pennil CC, Garcia R, Gross J, Sakai W, et al. Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. J Clin Oncol Off J Am Soc Clin Oncol. 2011 Aug 1;29(22):3008–15.

136. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, et al. Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature. 2008 Feb 28;451(7182):1116–20.

137. Vollebergh MA, Lips EH, Nederlof PM, Wessels LFA, Schmidt MK, van Beers EH, et al. An aCGH classifier derived from BRCA1-mutated breast cancer and benefit of high-dose platinum-based chemotherapy in HER2-negative breast cancer patients. Ann Oncol Off J Eur Soc Med Oncol. 2011 Jul;22(7):1561–70.

138. Vollebergh MA, Lips EH, Nederlof PM, Wessels LFA, Wesseling J, Vd Vijver MJ, et al. Genomic patterns resembling BRCA1- and BRCA2-mutated breast cancers predict benefit of intensified carboplatin-based chemotherapy. Breast Cancer Res BCR. 2014 May 15;16(3):R47.

139. Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilas E, Francoeur N, et al. Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2010 Aug 1;28(22):3555–61.

140. Mulligan JM, Hill LA, Deharo S, Irwin G, Boyle D, Keating KE, et al. Identification and validation of an anthracycline/cyclophosphamide-based chemotherapy response assay in breast cancer. J Natl Cancer Inst. 2014 Jan;106(1):djt335.

141. Maxwell KN, Wubbenhorst B, Wenz BM, De Sloover D, Pluta J, Emery L, et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun. 2017 Aug 22;8(1):319.

VII. Annexes

Annexe 1: "Current gene panels account for nearly all homologous recombination repair-associated multiple-case breast cancer families"

npj | Breast Cancer

www.nature.com/npjbcancer

BRIEF COMMUNICATION OPEN

Check for updates

Current gene panels account for nearly all homologous recombination repair-associated multiple-case breast cancer families

Thibaut S. Matis[1,2,3], Nadia Zayed[1,2], Bouchra Labraki[4], Manon de Ladurantaye[5], Théophane A. Matis[6], José Camacho Valenzuela[1,2], Nancy Hamel[2], Adrienne Atayan[1], Barbara Rivera[7,8,9], Yuval Tabach[10], Patricia N. Tonin[1,2,11], Alexandre Orthwein[7,8], Anne-Marie Mes-Masson[5,12], Zaki El Haffaf[4], William D. Foulkes[1,2,7,8]✉ and Paz Polak[13]✉

It was hypothesized that variants in underexplored homologous recombination repair (HR) genes could explain unsolved multiple-case breast cancer (BC) families. We investigated HR deficiency (HRD)-associated mutational signatures and second hits in tumor DNA from familial BC cases. No candidates genes were associated with HRD in 38 probands previously tested negative with gene panels. We conclude it is unlikely that unknown HRD-associated genes explain a large fraction of unsolved familial BC.

npj Breast Cancer (2021)7:109 ; https://doi.org/10.1038/s41523-021-00315-8

Many multiple-case BC families remain unexplained by a pathogenic variant, despite routine clinical testing of an increasing number of recognized BC susceptibility genes (BCSGs). These cases pose a major clinical challenge for disease management of patients and cancer prevention.

To identify germline pathogenic variants (GPVs) in patients, Next-Generation Sequencing (NGS) of blood-derived DNA is used, often through pan-cancer panels containing up to 25 genes. The three major BCSGs – BRCA1, BRCA2, and PALB2 – are involved in homology-directed DNA repair (HR). Several other, less well-studied genes, also involved in DNA repair and all variably associated with increased risk for hereditary breast and ovarian cancer (HBOC), are also included on some panels. As PARP inhibitors have been approved for treating cancers with HR defects, it is important to identify all possible HR pathway genes reliably implicated in BC susceptibility. To this end, many clinical BC susceptibility genes (BCSGs) testing panels include putative BC-related HR pathway genes, but their candidacy remains unproven. This can lead to inappropriate use of expensive therapies.

We and others showed that tumor sequencing is a powerful tool to find genes or variants that cause HRD breast cancer. In BC cases due to known BCSGs, the inherited pathogenic variant confers genetic susceptibility, but tumorigenesis is usually driven by a somatic inactivating "second-hit" event resulting in loss of gene function, often via loss of heterozygosity (LOH) through deletion of the wild-type allele. In this tumor suppressor model of BCSG function, as a nascent tumor cell loses the ability to repair DNA, lesions accumulate, including in genes favoring tumor progression[1]. Tumor genomes of patients with *BRCA1/2* inactivation were found to be enriched in copy number losses and rearrangements. This observation led to the development of an HRD index based on these somatic genomic abnormalities that can be used to detect HRD implicated tumors rather than on the specific mutational spectrum at the nucleotide level of the altered the DNA sequence[2–5]. In addition, in tumors with anomalies in HR repair, single base-pair changes have been shown to accumulate in the tumor genome, giving rise to what is referred to as mutational signatures. We showed that using sequencing of paired tumor and normal tissues we can detect these HRD-associated mutational signatures as well as second hits[6,7]. COSMIC tumor mutational signature 3 (Sig 3) has been associated with homologous repair deficiency (HRD)[8]. Monoallelic inactivation of *BRCA1/2* does not give rise to Sig 3, whereas Sig 3 is seen in tumors with biallelically inactivated *BRCA1/2*.

The mutational signature analysis helped to define the role of various HR genes directly from case-only data by association with Sig 3 levels. We showed that inactivation of *RAD51C*, *RAD51D*, *PALB2*, and *BARD1* are all associated with elevated Sig 3 levels, indicating that GPVs in these genes lead to *BRCA*-like/HRD cancers. In contrast, GPVs in other established and candidate BCSG genes such as *ATM*, *ATR*, *BRIP1*, *MRE11*, *NBN*, and *RAD50* do not lead to BRCA-like tumors[6]. Moreover, we developed a framework to reclassify germline *BRCA1/2* variants of unknown significance (VUS) as benign or pathogenic variants in the context of HRD in tumors[6]. A variant was classified as a pathogenic or disease-causing variant if the tumor that harbored it showed a second hit and Sig 3. This approach had 100% sensitivity and 98.5 % specificity in classifying 142 known benign and pathogenic missense mutations in *BRCA1/2*. Thus, Sig 3 levels can be used to identify new HR genes and new pathogenic variants in genes that lead to BRCA-like tumors.

In this study, we designed a pipeline encompassing variant calling and detection of second events involving somatic variants, combined with mutational signature analysis for HRD detection.

[1]Department of Human Genetics, McGill University, Montreal, QC, Canada. [2]Cancer Research Program, Centre for Translational Biology, The Research Institute of the McGill University Health Centre, Montreal, QC, Canada. [3]Cancer Genetics Unit, Institut Bergonié, Bordeaux, France. [4]Divison de Médecine Génétique, CRCHUM, Montreal, QC, Canada. [5]Centre de recherche du Centre hospitalier de l'Université de Montréal and Institut du cancer de Montréal, Montreal, Quebec, Canada. [6]Ecole Supérieur de Génie Informatique, Paris, France. [7]Gerald Bronfman Department of Oncology, McGill University, Montreal, QC, Canada. [8]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada. [9]IDIBELL Bellvitge Biomedical Research Institute, Barcelona, Spain. [10]Department of Developmental Biology and Cancer Research, Institute for Medical Research-Israel-Canada, Hebrew University of Jerusalem, Jerusalem, Israel. [11]Department of Medicine, McGill University, Montréal, QC, Canada. [12]Department of Medicine, Université de Montréal, Montreal, QC, Canada. [13]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ✉email: william.foulkes@mcgill.ca; paz.polak@mssm.edu

Published in partnership with the Breast Cancer Research Foundation

npj nature partner journals BCRF

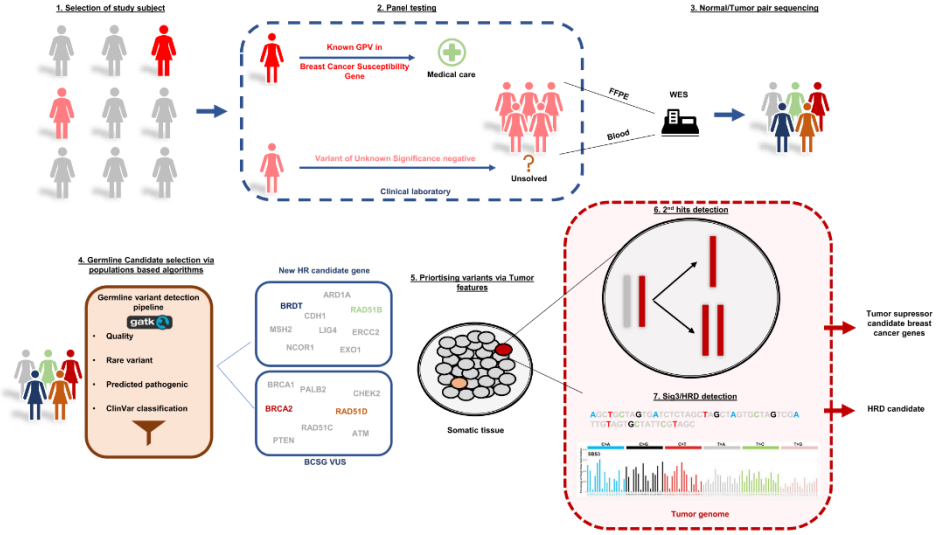


Fig. 1 Study workflow. (1) all eligible women were selected based on their personal and family history of cancer (red and pink women) and (2) received prior clinical screening. (3) unsolved cases were selected for WES through normal/tumor pair samples from blood lymphocytes and FFPE tissues. (4) sequencing data were analyzed through a germline candidate variant detection algorithm focused on homology-directed repair candidates following a classical BCSG model. (5) tumor landscapes were evaluated through a variant calling pipeline designed to minimize artifacts and a second-hit detection strategy based on the biological mechanism (i.e. defects in the hr pathway). (6) candidate BCSGs were re-evaluated in the context of a patient harboring a germline variant with an HRD tumor and second hit.

The pipeline was optimized for archival formalin-fixed paraffin-embedded (FFPE) tumor tissues. We studied 38 unsolved BC patients with either early-onset BC or a strong family history of BC with no identified GPVs in known BCSGs (Supplementary Table 1). We tested the hypothesis that their tumors are linked to HRD due to alterations in genes that have not yet been definitively found associated with heritable risk to BC (Fig. 1).

To identify possible genetic causes that underly BC in our cases, we compiled a list of candidate genes implicated in HR which are not on current genetic testing panels (see methods). We analyzed whole-exome sequencing data (WES) of tumor/normal pairs to look for potentially pathogenic rare variants and to search for evidence of a second hit in the tumor based on a tumor suppressor gene model as well as for evidence of an HRD signature[6] (Fig. 1).

Three identified variants were considered as pathogenic by ClinVar: two variants were in the *FANCA* and *GJB2* genes, but neither showed a second hit. The third was a nonsense variant in *LIG4* that showed LOH in the tumor. We observed LOH of the WT allele for 18 additional germline variants. We found five VUSs and three variants with Conflicting Interpretation of pathogenicity (CIOP) in *APC, ATM*, *BRCA2*, *CHEK2*, *FLCN*, and *RAD51D*; and the remaining variants were not classified in ClinVar (Fig. 2, Supplementary Fig. 1). No second somatic LOF alteration was found in combination with a germline candidate variant. All variants are summarized in Tables 1, 2.

To identify patients with HRD-related tumors, we used two HRD classifiers tools. SigMA is tailored for WES tumor data and is used to identify Sig 3 (Fig. 2, Supplementary Fig. 2). We also used ScarHRD, which is based on evaluating genomics scars due to HRD in tumor DNA and has been recently adapted for use with WES data. We identified 7 HRD tumors based on SigMA, and 12 based on ScarHRD (Supplementary Fig. 3). All 7 cases exhibiting an HRD signature by SigMA analysis were those also classified as having genomic scars consistent with HR defects by ScarHRD. Given that WES is not ideally suited to detect large genomic events, we considered the 7 tumors (18%) that were called positive by both tools for further analysis (Fig. 2). This percentage is within the range of detection reported by other HRD studies[7,9–11]. The only HRD tumor with a germline BCSG candidate variant was from a BC patient with the missense variant *RAD51D* p.S207L (case C10758). Using copy number analysis, we were able to identify duplication of the germline missense variant along with a *trans* allele loss, resulting in copy neutral LOH (CN-LOH) event at the gene locus as the second hit. We found a somatic deletion at a donor splice-site in *RAD51C* associated with duplication-LOH in another tumor with HRD (case JGH1), without germline events (Fig. 2). No other HRD tumor was found to be associated with our candidate genes either somatically or in the germline.

Variants in known BC drivers[12] were found in 84% of tumors (Fig. 2, Supplementary Fig. 4). Of these, *TP53, GATA3, and CBFB* were considered as a driver (see methods). Three patients harbored a *PIK3CA* hotspot mutation. We also detected somatic alterations in three BC genes (*PTEN, PALB2*, and *CDH1*) in 4 other patients. Copy number profiles were similar to observations from previous BC cohort studies[12,13].

Our pipeline worked well in calling variants in archival FFPE tumors, which are more accessible than fresh tumor material. Except for the *RAD51D* variant, which has a conflicting interpretation of pathogenicity (https://www.ncbi.nlm.nih.gov/clinvar/variation/142102/), we did not find supporting evidence to reclassify any ClinVar-described variants found in our HRD tumors as pathogenic. The fact that we were able to independently identify p.S207L *RAD51D* variant as a candidate pathogenic variant using our pipeline demonstrates the robustness of the genomic tools used in this study to reassess VUS. Although we sequenced a highly selected population of BC patients from multiple-case BC families, we did not find any other plausible GPVs in genes beyond

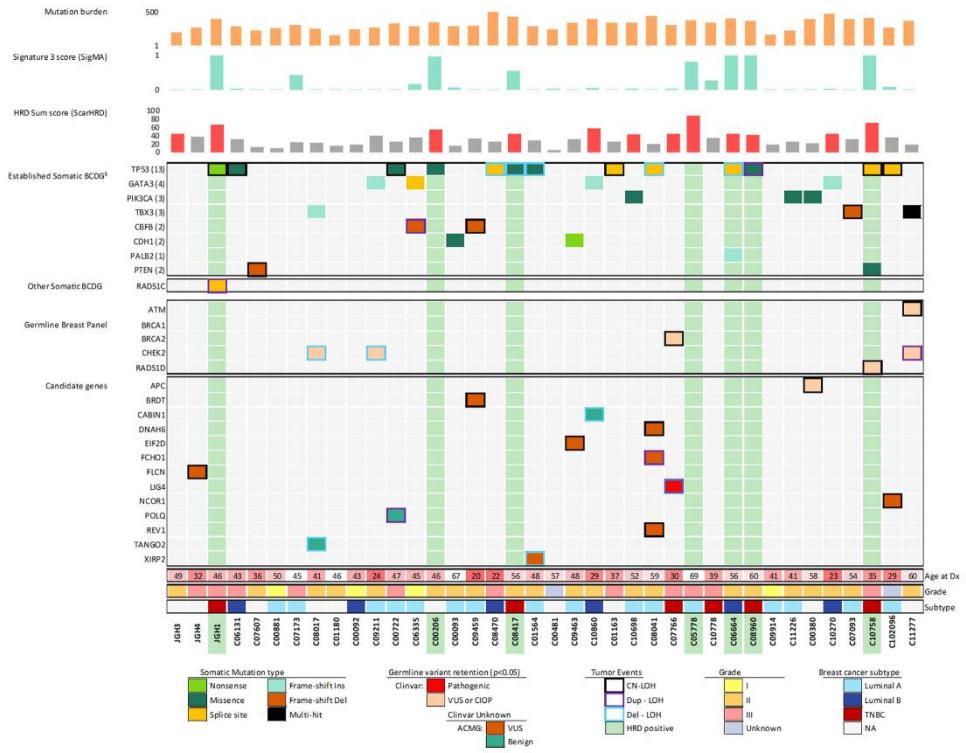


Fig. 2 Co-mutation plot of germline and somatic landscape after HRD detection. The HRD status evaluated by SigMA (that estimates the presence of Signature 3) is shown in the second row from the top, and columns with HRD based on SigMA are highlighted in vertical green columns. ScarHRD, shown in row 3, calculates the HRD index and cases with HRD determined by this method that has scores great than 41 are represented by red bars. The germline and somatic mutation status of candidate genes across patients are shown for four classes of genes (rows 4–7). BCDG: Breast Cancer Driver Gene; CN-LOH: copy neutral LOH; Dup-LOH: Duplication with LOH; Del-LOH: hemizygous deletion with LOH.

those currently included in the clinical gene testing panels, which is in line with recent studies[14]. We focused on HRD as a causal mechanism in BC, but our approach could be of use to identify germline drivers in other tumor types with mutational signatures that are associated with defects in DNA repair pathways, such as mismatch repair, hyper-mutated tumors, and base excision repair.

There are a number of caveats to our study. Our inability to detect potentially pathogenic variants in recently confirmed BCSGs (e.g. *RAD51C* and *BARD1*[10,15]) as well in largely unexplored HR genes could be due to the small sample size of our study. As WES is not ideally suited to identify large structural rearrangements associated with HR defects we could have missed tumor harboring anomalies due to HRD. Also, it is possible that undiscovered BCSGs do not operate via the HR pathway, and in this case our HRD-based framework will not detect variants in these genes. Finally, we cannot exclude the possibility that variants were missed for technical reasons given the age of some of the FFPE material, as one-quarter of our cases were diagnosed more than 10 years prior to analysis, which likely affected the quality of DNA for analysis. This limitation would not be as relevant for somatic testing of freshly processed tissues in a clinical context. Importantly, our study focused on WES data, so we could not assess other molecular mechanisms affecting gene function as potential contributors, such as complex structural variations[9] or somatic methylation events[16].

Nones et al.[17] took a similar approach using fresh frozen tumors and were able to show that all but one causally unexplained BCs with HRD were due to inactivation of *BRCA1*, *BRCA2*, or *PALB2*. Taken together with the results we present here, we can conclude that the tail of missing heritability due to HRD genes must be long and thin. It also means that existing gene testing panels are sufficient to identify carriers of clinically important BC genes using a normal/tumor pair sequencing-based approach. We must therefore look elsewhere to find the missing inherited susceptibility present in these families.

METHODS

Samples selection

This study called "Genome-wide approaches in hereditary cancer families", study number A08-M61-09B, was created in 2011 and approved by the McGill University Research Ethics Board. Participants to this study were consented at two Montreal sites. Some participants were consented directly into the scientific study at McGill University and its affiliated hospitals. Others were recruited from a Montreal-based biobank entitled "Banque d'échantillons biologiques et de données (cliniques et biologiques) associées à des fins de recherche sur les cancers métastatiques du

Table 1. List of missenses variants retained in the tumor by LOH.

Gene	Patient ID	NM number	DNA changes	Amino Acid changes	Variant consequences	Gene category	ClinVar classification	ACMG classification	Variant retention type	Germline REF read counts	Germline ALT read counts	Tumor REF read counts	Tumor ALT read counts	*p* value
APC	C00380	NM_001127510	c.8275C>T	p.R2759C	Missense	CSG	VUS	–	CN-LOH	80	64	134	177	0.013197
ATM	C11277	NM_000051	c.1538A>G	p.Q513R	Missense	BCSG	VUS	–	CN-LOH	191	142	8	24	0.000446
BRCA2	C07766	NM_000059	c.9744G>A	p.M3248I	Missense	BCSG	VUS	–	CN-LOH	114	97	13	52	0.000669
BRDT	C09459	NM_001242805	c.272C>T	p.A91V	Missense	C-HRG	Unknown	VUS	CN-LOH	42	35	12	38	0.00129
CABIN1	C10860	NM_012295	c.2446C>T	p.L816F	Missense	C-HRG	Unknown	Likely benign	DEL-LOH	112	107	32	55	0.032096
CHEK2	C08017	NM_001005735	c.628G>A	p.G210R	Missense	BCSG	ClOP	–	DEL-LOH	211	155	22	103	<0.00001
CHEK2	C09211	NM_001005735	c.1165C>T	p.R389C	Missense	BCSG	VUS	–	DEL-LOH	114	128	36	67	0.037132
CHEK2	C11277	NM_001005735	c.1465A>G	p.N489D	Missense	BCSG	VUS	–	DUP-LOH	192	180	23	49	0.002237
EIF2D	C09463	NM_006893	c.1538A>G	p.Y513C	Missense	C-HRG	Unknown	VUS	CN-LOH	35	36	136	406	0.000016
FCHO1	C07766	NM_001161357	c.557G>A	p.R186Q	Missense	C-HRG	Unknown	VUS	DUP-LOH	155	148	31	87	<0.00001
FLCN	JGH4	NM_144997	c.715C>T	p.R239C	Missense	CSG	ClOP	–	CN-LOH	109	120	44	185	<0.00001
NCOR1	C102096	NM_006311	c.2794C>T	p.H932Y	Missense	C-HRG	Unknown	VUS	CN-LOH	138	101	34	99	<0.00001
POLQ	C00722	NM_199420	c.1981T>G	p.W661G	Missense	C-HRG	Unknown	Likely Benign	DUP-LOH	26	28	7	21	0.042667
RAD51D	C10758	NM_002878	c.620C>T	p.S207L	Missense	BCSG	ClOP	–	CN-LOH	129	120	13	132	<0.00001
REV1	C08041	NM_001321460	c.2017G>A	p.D673N	Missense	DNA Repair Gene	Unknown	VUS	CN-LOH	88	80	60	93	0.01811

For the three ClOP variants, *FLCN* c.715C>T is trending towards benign https://www.ncbi.nlm.nih.gov/clinvar/variation/134427/ and is not a plausible BCSG. *RAD51D* c.620 C>T is discussed above and is recorded as likely pathogenic by most laboratories https://www.ncbi.nlm.nih.gov/clinvar/variation/142102/. *CHEK2* c.628 G>A (also known as NM_007194.4(CHEK2):c.499 G>A (p.Gly167Arg)) is reported as likely pathogenic by nearly all laboratories on ClinVar https://www.ncbi.nlm.nih.gov/clinvar/variation/142524/ and has been reported to result in a severe cancer phenotype in the homozygous state[40]. While the presence of LOH in the case reported here could contribute to its final classification on ClinVar as Pathogenic, biallelic *CHEK2* mutations are not associated with any known mutational signature 3 and therefore our approach cannot be used to reclassify this or any other *CHEK2* variant.
BCSG breast cancer susceptibility genes, *CSG* cancer susceptibility genes, *C-HRG* candidate homologous recombination genes, *DEL-LOH* LOH by deletion of the WT allele, *DUP-LOH* LOH by multiplication of the alternate variant and loss of the WT allele, *CN-LOH* LOH by copy neutral event, *VUS* variant of uncertain significance, *ClOP* variant with conflicting interpretations of pathogenicity.

Table 2. List of truncating variants retained in the tumor by LOH.

Gene	Patient ID	NM number	DNA changes	Amino Acid changes	Variant consequences	Frequency of LoF	Gene category	ClinVar classification	ACMG classification	Variant retention type	Germline REF read counts	Germline ALT read counts	Tumor REF read counts	Tumor ALT read counts	p value
DNAH6	C08041	NM_001370	c.6184delG	p.D2062Mfs*19	frameshift del	1.08%	C-HRG	Unknown	Likely Pathogenic	DEL-LOH	79	34	35	48	0.000181
LIG4	C07766	NM_206937	c.2440C>T	p.R814X	nonsense	0.16%	DNA Repair Gene	Pathogenic	-	DEL-LOH	148	83	29	40	0.001089
TANGO2	C08017	NM_001322149	c.158C>G	p.S53X	nonsense	1.22%	C-HRG	Unknown	Benign	CN-LOH	194	187	49	204	<0.00001
XRP2	C01564	NM_152381	c.8352delG	p.L2784Ffs*43	frameshift del	1.44%	C-HRG	Unknown	Likely Pathogenic	DEL-LOH	69	56	18	33	0.016566

LoF frequencies (LF) for an x gene were calculated through $LF(x) = 2 \times \left(\frac{S(x)}{m(x)}\right)$ where S is the allelic count of LoF variant, m the median of all allele number in gnomAD v2.1.1.

sein et de l'ovaire" previously created in 2000 and approved by the CHUM Institutional Review Board, approval number BD 04.002. We identified over 70 candidate families eligible for this study. All had early-onset BC or a strong family history of breast cancer (median Manchester score 24, range 10-59). None carried a *BRCA1/2* or *PALB2* pathogenic or likely pathogenic variant on clinical genetic testing. We reviewed approximately 130 pathology blocks to retrieve available BC tumor material that would be suitable for analysis. Blood samples were only obtained from patients where tumor material was successfully retrieved and processed. Of 59 FFPE tumors with extracted DNA, 20 samples failed sequencing due either to poor quality or insufficient quantity of DNA. WES was thus performed on 39 normal/tumor pairs of women with BC diagnosed between 1995 and 2017, but we removed one because of unresolved contamination issues. The interval between diagnosis and sequencing ranged from a few months to 22 years (median 3 years).

Normal/tumor WES sequencing

Genomic DNA was extracted from the patient's peripheral blood lymphocytes and tumor using published commercial protocols and kits optimized in our lab (Maxwell RSC Buffy Coat DNA/Tissue DNA kits, Promega). Library preparation, exome capture, and sequencing were conducted over a period of 2 years, as material was obtained, reflecting repeat assays for DNA samples that failed. A Panel of Normals (PoN) comprised of 47 normal tissue samples, all sequenced using the same protocol as we applied to tumor tissues, was used to assist in removing recurrent technical artifacts due to formalin fixation. Genomic DNA libraries were generated using one of the following: Nextera Rapid Capture Exome Target (Illumina, San Diego, California), Sure Select Human All Exon V6, or Sure Select Human All Exon V7 (Agilent). DNA samples were then sequenced using Illumina sequencers and appropriate protocols available at the time of sequencing: Illumina HiSeq 2000 as 100 bp paired-end reads, Illumina HiSeq 2500 as 125 bp paired-end reads, and Illumina HiSeq 4000 as 100 bp paired-end reads (Illumina, San Diego, California).

Germline variant calling and filtering

Exome sequencing data analysis was performed using the pipeline we developed for this project. Briefly, BWA (v. 0.5.9) and the Genome Analysis Toolkit (GATK)[18] were used to align the sequenced reads to the reference genome (UCSC hg19) and to perform local realignment of reads around small insertions and deletions (indels). PCR duplicate reads were marked using Picard (https://github.com/broadinstitute/picard). The coverage of consensus coding sequence (CCDS) bases was calculated by GATK, resulting in a mean coverage of 141.3X (ranging from 26.9X to 250.6X) and 97.3% of CCDS bases were covered by at least 10 reads. Germline variants were called individually for each sample using HaplotypeCaller from GATK and annotated by wAnnovar[19]. From the generated list of germline variants, variants were filtered out if they fulfilled any one of the following quality criteria: (i) genomic position of variant covered by <5 reads; (ii) <2 reads support the alternative variant; (iii); allelic fraction between < 0.35 or >0.5; (iv) and mapping quality < 30. We then focused on four categories of genes (n = 864): 20 BCSG frequently used in clinical testing, and previously tested in our HBC cases; 594 "Candidate HR genes" list (C-HRGs) purported to have co-evolved[20] or known to interact in HR (unpublished data from Dr. Alexander Orthwein); 103 known DNA repair genes; and 147 other known Cancer Susceptibility Genes (CSGs)[21] (Supplementary Table 2). Only variants having an allele frequency <0.01 based on allele frequencies in GnomAD[22] and 1000 Genomes project[23] were retained for further investigation.

Somatic variant calling and filtering

Somatic variants were called individually through normal/tumor paired caller using Mutect2[24] for SNV and indel. The PoN was used as well as germline population resources according to GATK's best practice. Variants were filtered out if they (i) displayed strand bias, to remove artifacts caused by the formalin fixation process; (ii) were germline in origin; (iii) were due to technical artifacts revealed by comparison to PoN, and (iv) had a depth < 10 reads. Variants were next annotated with VEP[25].

Somatic Copy Number Analysis and LOH detection

With GATK CNA, read coverage counts across the target were collected. We created a separate copy number variation (CNV) PoN to reduce read coverage data noise to obtain denoising copy ratios against the PoN. Then

we collected counts of reference versus alternate alleles for allelic imbalance. We used GISTIC[26] to assess recurrent focal amplifications and deletions in tumors. We used FACETS[27] and Sequenza[28] to infer the integer copy number of each pair of alleles as well as LOH. Hemizygous deletions were defined as the loss of one allele in the tumor. Homozygous deletions were defined as the loss of both alleles in the tumor. Amplification was defined as a copy number status ≥ 6 per tumor for *ERBB2*, *MYC*, and *ZNF703*.

Variant visualization

The Integrative Genomics Viewer[29] was used for the manual examination and visualization of all germline candidate variants, and somatic events.

Somatic cancer driver gene identification

A combination of several tools was used to infer drivers of oncogenesis based on statistical methods through q-score to identify genes that were mutated more often than expected by chance: MutSigCV[30], OncodriveFML[31], MutPanning[32], and FishHook[33]. Genes that were found mutated more than once with a q-score < 0.05 were considered as cancer driver genes.

Mutational signature and homologous recombination repair deficiency detection

Different tools were used to assess the mutational signature from normal-tumor paired WES VCF data: MutationalPatterns[34], Maftools[35], DeconstructSig[36] for tumor signature clustering and the mutational signature contribution in each tumor. Different approaches were used: a de novo mutational signature extraction using a non-negative matrix factorization (NMF) and a COSMIC signature similarity using cosine similarity. We used SigMA[37] to detect the mutational signature associated with HR defect designed for WES. We also used ScarHRD[38] to detect HRD genomic scars—genomic abnormalities characteristic of HRD, which is based on the sum of LOH score, Large Scale Transition (LST) score, and Telomeric Allelic Imbalance (TAI) score. A score of ≥ 42 classifies tumors as HRD (see Supplementary Fig. 3).

Candidate variant selection algorithm

A list of candidate pathogenic variants was generated using our variant detection algorithm (Supplementary Fig. 5). ClinVar annotations were used to classify variants already reported, and when no ClinVar data was available, we used the ACMG criteria. Variants classified as likely benign or benign by ClinVar were filtered out. The master list of variants was comprised of truncating, missense, or splice-site alterations found in gene candidates.

For non BSCG, missense and splice-site variants predicted to be deleterious by at least 4 of 8 bioinformatic prediction tools (SIFT, PolyPhen2, MutationTaster, FATHM, Provean, MetaSVM, MetaLR, and CADD) were considered for further analysis. For known BSCG genes, we considered all variants regardless of scores as these can be evaluated by expertise. Truncating variants were classified separately. The candidate list was comprised of 231 variants: 7 missense VUSs or variants with conflicting interpretations of pathogenicity in known BCSGs, 46 truncating variants, and 178 predicted pathogenic variants in C-HRGs, DNA repair genes, and in CSGs. From this list of candidate genes, we searched for evidence of allelic imbalance at each gene locus in the tumor using statistical tests (Chi-squared test) and filtered out variants where the p-value was >0.05 and confirmed with FACET results that there was a deletion of one haplotype (see above). We also looked for somatic second-hit point mutations. Our final list of candidate variants was comprised of genes that showed evidence of a possible second mutational hit in the tumor.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data generated and analyzed during this study are described in the following data record: https://doi.org/10.6084/m9.figshare.14802954[39]. Tumor and germline datasets generated during the current study are not publicly available as the consent provided by study participants did not include a provision for widespread disclosure. However, direct requests can be made to the corresponding authors for access. The following files are openly available as part of the figshare data record: 'Master_list_variant.xlsx' (underlying Table 1), 'Genes assessed initially.xlsx' (underlying Supplementary Table 1), 'Candidate Genes list.xlsx' (underlying Supplementary Table 2). All other data are housed on institutional storage and are not openly available in order to protect patient privacy as informed consent to share participant-level data was not obtained prior to or during data collection. However, these data can be requested from Dr Foulkes. The files are: 'all germline_sample.vcf', 'data.tumor.maf', 'all tumor_sample.vcf', 'all sample.out.gzsmall.seqz.gz'.

CODE AVAILABILITY

The bioinformatics analyses were performed using open-source software, including Burrows–Wheeler alignment tool (v. 0.5.9), GATK (v.4.1.6.0), samtools (v.1.10), bcftools (v.1.10), htslib (v.1.10), Picard (v.2.24.2), vcf2maf (v.1.6.17), R (v.4.0.3), CNV_facets (v.0.16.0), Maftools (v.0.9.30), FishHook (v.1.0), Bioconductor (v.3.12), MutationalPatterns (v.3.0.1), deconstructSigs (v.1.0), SigMA (v.1.0), Sequenza (v.3.0.0) and ScarHRD (v.1.0).

Received: 12 February 2021; Accepted: 26 July 2021;
Published online: 25 August 2021

REFERENCES

1. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2011).
2. Abkevich, V. et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br. J. Cancer* **107**, 1776–1782 (2012).
3. Popova, T. et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **72**, 5454–5462 (2012).
4. Birkbak, N. J. et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov.* **2**, 366–375 (2012).
5. Telli, M. L. et al. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. *Clin. Cancer Res.* **22**, 3764–3773 (2016).
6. Polak, P. et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genet.* **49**, 1476–1486 (2017).
7. Davies, H. et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational-signatures. *Nat. Med.* **23**, 517–525 (2017).
8. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
9. Nguyen, L., W. M. Martens, J., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**, 5584 (2020).
10. Riaz, N. et al. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nat. Commun.* **8**, 857 (2017).
11. Wang, Y. K. et al. Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* **49**, 856–865 (2017).
12. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
13. Sircoulomb, F. et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol. Med.* **3**, 153–166 (2011).
14. Narod, S. A. Which genes for hereditary breast cancer? *N. Engl. J. Med.* **384**, 471–473 (2021).
15. Breast Cancer Association Consortium. et al. Breast cancer risk genes—association analysis in more than 113,000 women. *N. Engl. J. Med.* **384**, 428–439 (2021).
16. Brianese, R. C. et al. BRCA1 deficiency is a recurrent event in early-onset triple-negative breast cancer: a comprehensive analysis of germline mutations and somatic promoter methylation. *Breast Cancer Res. Treat.* **167**, 803–814 (2018).
17. Nones, K. et al. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. *Ann. Oncol.* **30**, 1071–1079 (2019).
18. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
19. Chang, X. & Wang, K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J. Med. Genet.* **49**, 433–436 (2012).
20. Sherill-Rofe, D. et al. Mapping global and local coevolution across 600 species to identify novel homologous recombination repair genes. *Genome Res.* **29**, 439–448 (2019).
21. Huang, K. et al. Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355–370.e14 (2018).
22. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

23. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
24. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
25. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
26. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
27. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
28. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
29. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
30. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
31. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **17**, 128 (2016).
32. Dietlein, F. et al. Identification of cancer driver genes based on nucleotide context. *Nat. Genet.* **52**, 208–218 (2020).
33. Imielinski, M., Guo, G. & Meyerson, M. Insertions and deletions target lineage-defining genes in human cancers. *Cell* **168**, 460–472.e14 (2017).
34. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
35. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
36. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
37. Gulhan, D. C., Lee, J. J.-K., Melloni, G. E. M., Cortés-Ciriano, I. & Park, P. J. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nat. Genet.* **51**, 912–919 (2019).
38. Sztupinszki, Z. et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* **4**, 16 (2018).
39. Matis, T. S. et al. Metadata record for the article: Current gene panels account for nearly all homologous recombination repair-associated multiple-case breast cancer families. figshare https://doi.org/10.6084/m9.figshare.14802954 (2021).
40. Paperna, T. et al. Homozygosity for CHEK2 p.Gly167Arg leads to a unique cancer syndrome with multiple complex chromosomal translocations in peripheral blood karyotype. *J. Med. Genet.* **57**, 500–504 (2020).

ACKNOWLEDGEMENTS

This study, including the acquisition of archived samples, was funded by a Quebec Breast Cancer Foundation (QBCF) grant to W.D.F. and P.T. Tumor banking was supported in part by the Banque de tissus et de données of the Réseau de recherche sur le cancer of the Fonds de recherche du Québec Santé (FRQS) affiliated with the Canadian Tumor Repository Network (CTRNet). Y.T. is supported by Israel Science Foundation grant. A.O. is supported by CCSRI Innovation Grant and Canada Research Chair Tier 2 in Genome Stability and Hematological Malignancies. P.P. is supported by the Jimmy V foundation scholar award.

AUTHOR CONTRIBUTIONS

T.S.M. developed the pipeline, analyzed the data, and drafted the manuscript. N.Z., B.L., Z.E.H., N.H., A.M.M.M., P.N.T., and W.D.F. provided clinical data or samples. N.Z. and M.D. L. prepared the samples. T.A.M. and J.C.V. contributed to the pipeline development. A. A. calculated the probability of pathogenic variants in *BRCA1/2* and related genes. B.R., N.H., P.P., Y.T., A.O., and W.D.F. validated the data. T.S.M., B.R., N.H., P.P., and W.D.F. wrote the manuscript. All authors reviewed and edited the manuscript.

COMPETING INTERESTS

The authors declare no competing interests

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41523-021-00315-8.

Correspondence and requests for materials should be addressed to W.D.F. or P.P.

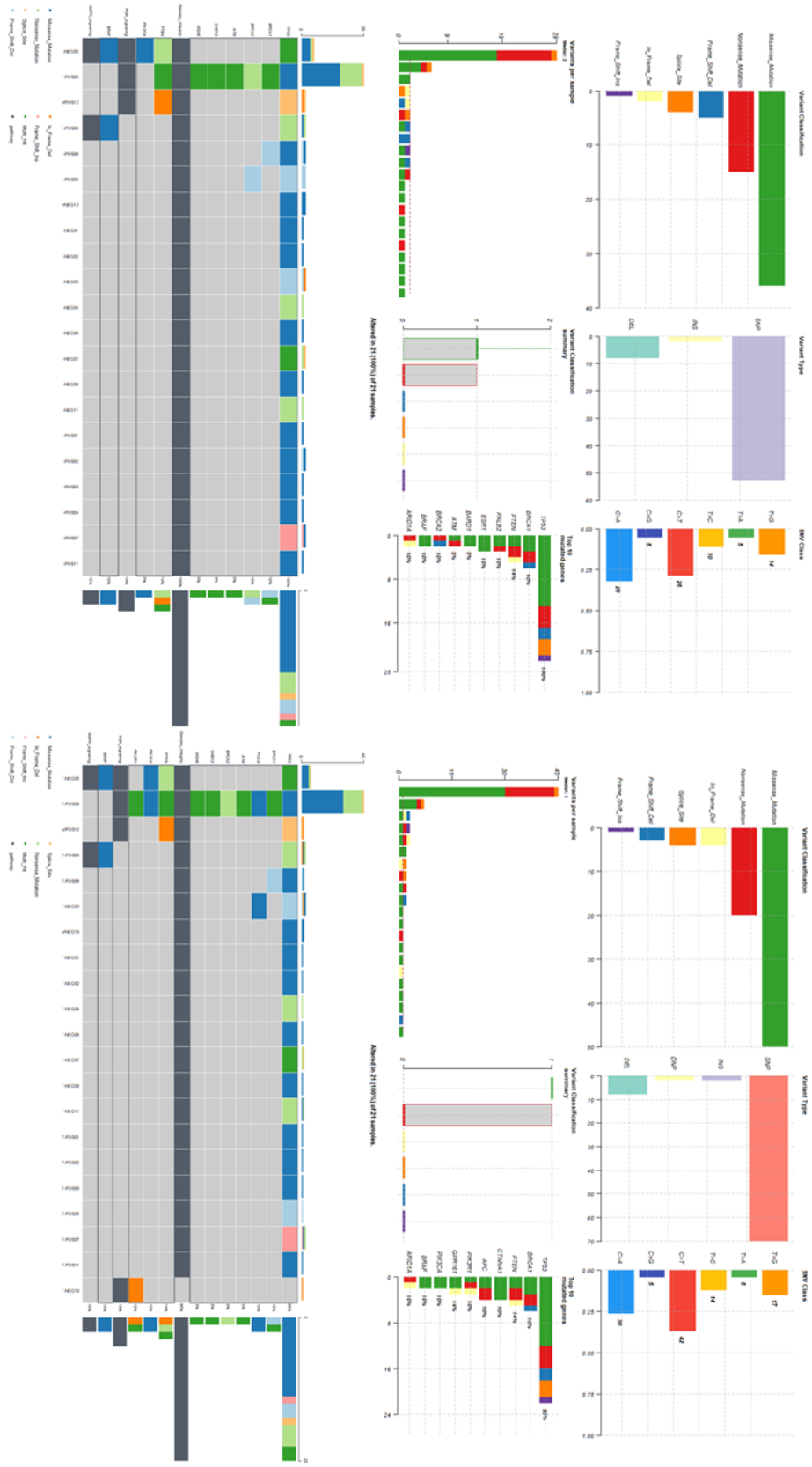
Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

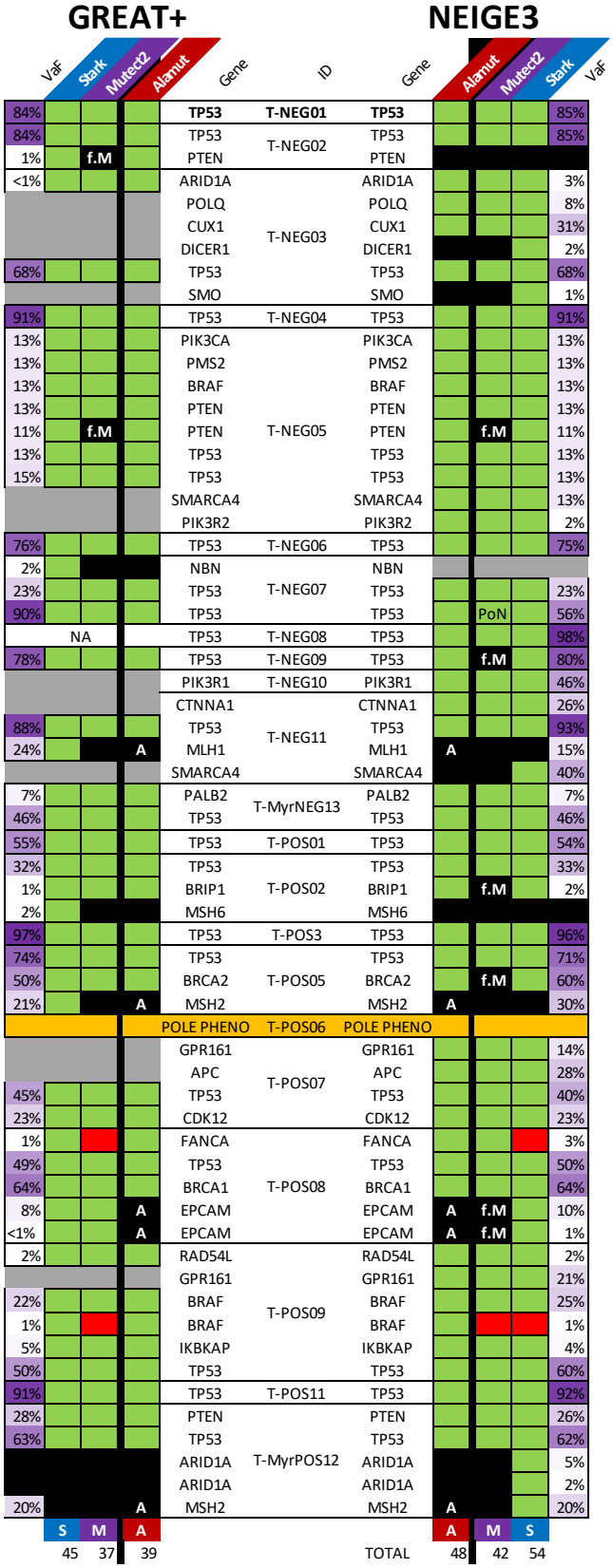
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

Annexe 2 : Paysages mutationnelles des 22 HGSC selon le panel GREAT+ (en haut) et NEIGE3 (en bas).



Annexe 3 : Comparaison des variants appelés par Mutect2 et STARK (et vérifiés sur Alamut) selon l'utilisation du panel GREAT+ et NEIGE3. Un carré vert signifie qu'un variant présent sur les BAM (Alamut) est retrouvé. Un carré rouge qu'un variant n'est pas retrouvé. Un carré noir qu'un variant absent n'est pas retrouvé. f.M : filtré par Mutect2 ; A : Artéfact ; PoN : variant filtré par PoN



Titre : **Approches bio-informatiques de caractérisation d'une signature moléculaire HRD sur des cohortes de patientes atteintes de cancer du sein ou des ovaires en tissus FFPE.**

Résumé :

Environ 25 à 50% des cancers du sein et de l'ovaire présenteraient un déficit dans la voie de réparation par recombinaison homologue (HRD). Les tumeurs HRD sont sensibles aux inhibiteurs de la poly-ADP ribose polymérase (PARPi) dont l'AMM a été récemment accordé dans les cancers du sein et de l'ovaire HRD. Ces tumeurs sont historiquement reliées à des altérations constitutionnelles ou somatiques, dans les gènes *BRCA1*, ou *BRCA2,* mais ne sont retrouvées que dans 60% des tumeurs HRD. Ce constat a fait émerger la caractérisation d'une signature moléculaire HRD comme test d'éligibilité aux PARPi plutôt que les tests ciblés sur les gènes *BRCA1/2.* La majorité des outils bio-informatiques permettant la détection de cette signature est développée à partir de tumeurs séquencées par génomes entier ou exomes (WES), et issues de tissus frais, difficilement extrapolable à la clinique. A l'aide d'un premier pipeline bio-informatique, nous avons évalué la détection d'une signature HRD sur des données de WES provenant de tissus FFPE par une analyse en normal/tumeur, puis nous l'avons extrapolée à des données issues de panels de gènes utilisés en routine. Ce pipeline est capable de détecter des tumeurs HRD issues d'échantillons FFPE par séquençage par WES. En revanche, l'utilisation de panels de gènes de routine, ne permet pas actuellement la détection des signatures mutationnelles associées à l'HRD.

Mots clés : HRD, PARP, pipeline bio-informatique, cancer du sein, cancer de l'ovaire, WES, panel de gènes

---

Title: **Bioinformatic approaches to establish an FFPE HRD molecular signature in a breast and ovarian cancer patient cohort.**

Abstract:

Approximately 25-50% of breast and ovarian cancer have a homologous recombination repair pathway deficiency (HRD). HRD tumors are sensitive to poly-ADP ribose polymerase inhibitors (PARPi), which have recently been approved for use in HRD breast and ovarian cancers. These tumors were originally associated with germline or somatic alterations in HR pathway genes, such as *BRCA1* or *BRCA2* found in 60% of HRD tumors. This observation has led to the emergence of HRD molecular signatures as an eligibility test for PARPi rather than *BRCA1/2* targeted testing. The majority of bioinformatics tools enabling the detection of such signatures have been developed from whole genome or exome (WES) sequencing from fresh frozen tumors, which is difficult to extrapolate in the clinical setting. Using a new bioinformatics pipeline, we first assessed the ability to detect the HRD signature on WES data from FFPE tissues using a normal/tumor analysis. Then, we extrapolated this pipeline with data from targeted panels used in diagnostic laboratories. This pipeline successfully detects HRD tumors from FFPE samples using WES sequencing, but not targeted panels.

Keywords : HRD, PARP, Bioinformatics pipeline, Breast cancer, Ovarian Cancer, WES, Targeted panel