

Identification de nouveaux gènes de prédisposition héréditaire au cancer du sein par génotypage tumoral et séquençage de nouvelle génération

Virginie Bubien

► To cite this version:

Virginie Bubien. Identification de nouveaux gènes de prédisposition héréditaire au cancer du sein par génotypage tumoral et séquençage de nouvelle génération. Génétique. Université de Bordeaux, 2016. Français. NNT : 2016BORD0393 . tel-01505203

HAL Id: tel-01505203 https://theses.hal.science/tel-01505203

Submitted on 11 Apr 2017 $\,$

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

Ecole doctorale Sciences de la Vie et de la Santé

Thèse pour l'obtention du Doctorat de l'Université de Bordeaux Spécialité : Génétique

Présentée et soutenue publiquement le 12 décembre 2016

Par Virginie BUBIEN

Née le 2 novembre 1979 à Guérande

Identification de nouveaux gènes de prédisposition héréditaire au cancer du sein par génotypage tumoral et séquençage de nouvelle génération

Thèse dirigée par le Docteur Michel LONGY

Membres du jury

M. le Professeur Benoît ARVEILER

Mme. le Docteur Nadine ANDRIEUX

M. le Professeur Stéphane BEZIEAU

M. le Professeur Cyril GOIZET

Président

Rapporteur

Rapporteur

Examinateur

Remerciements

Les travaux présentés dans cette thèse ont été réalisés au sein de l'Unité INSERM 1218 «Actions for oncogenesis understanding and Target Identification in Oncology » dirigée par le Professeur Pierre Soubeyran que je remercie pour m'avoir accueillie dans son laboratoire. J'adresse également mes remerciements au programme « Carte d'Identité des Tumeurs » de la Ligue contre le Cancer et à la société IntegraGen pour les analyses réalisées.

Pour l'honneur qu'ils me font en faisant partie de ce jury, je remercie profondément le Professeur Benoît Arveiler qui a accepté de le présider, le Docteur Nadine Andrieu et le Professeur Stéphane Bézieau qui ont bien voulu juger ce travail, et le Professeur Cyril Goizet qui a accepté de l'examiner.

Je remercie mon directeur de thèse, le Docteur Michel Longy, pour son professionnalisme, sa rigueur et les connaissances qu'il m'a transmises, pour le temps et l'attention qu'il m'a consacrés et pour le soutien et la confiance qu'il m'a accordés durant toutes ces années.

Je remercie également les Docteurs Gaëtan MacGrogan et Isabelle Soubeyran pour leur contribution ainsi que les techniciennes de leur unité qui ont participé à ce projet.

Je tiens à remercier l'ensemble des membres de l'équipe d'Oncogénétique de l'Institut Bergonié, pour leur accueil et leur enseignement. Merci Delfine et Bernadette pour les techniques réalisées. Merci Emmanuelle, Françoise et Nicolas pour votre générosité, votre soutien et vos nombreux conseils ! Merci Natalie pour ton anglais ! Merci Alice et Claire pour m'avoir fait rire souvent quand j'avais envie de pleurer. Merci à mes assistantes médicales adorées, Florence et Elisabeth, pour les mêmes raisons.

Mes remerciements vont aussi à mes amis, ma famille et ma belle-famille, pour leur amour et leur soutien durant toutes ces années!

Mes plus grands remerciements sont évidemment pour toi Mathieu et pour nos filles chéries.

Table des matières

Table des fig	gures					
Liste des tab	leaux					
Liste des An	Liste des Annexes					
Liste des abr	Liste des abréviations					
Résumé	Résumé					
Abstract						
1 INTRO	DUCTION12					
1.1 Epi	démiologie du cancer du sein12					
1.2 Étic	blogies du cancer du sein					
1.2.1	L'âge					
1.2.2	Le mode de vie					
1.2.3	Les maladies bénignes du sein14					
1.2.4	La densité mammaire					
1.2.5	L'exposition aux radiations ionisantes					
1.2.6	L'exposition aux hormones endogènes et exogènes16					
1.2.7	Les antécédents de cancer du sein17					
1.3 Cla	ssification des cancers du sein17					
1.3.1	Classification histologique, phénotypique et pronostique18					
1.3.2	Classification moléculaire des cancers du sein					
1.3.3 prédisp	Caractéristiques des cancers du sein survenant dans le cadre d'une osition héréditaire connue					
1.4 Pré	disposition héréditaire au cancer du sein					
1.4.1	Les formes familiales de cancer du sein et leur déterminisme mendélien					
1.4.2 cancer of	Les stratégies d'identification des facteurs génétiques de prédisposition au lu sein					
1.4.3	Les gènes de prédisposition héréditaire au cancer du sein					
2 OBJEC	TIFS DE TRAVAIL					
3 MATÉI	RIEL ET MÉTHODES					
3.1 Ma	tériel					
3.2 Mé	thodes					
3.2.1	Extraction d'ADN					
3.2.2	Génotypage tumoral par puces SNP haute résolution					
3.2.3 [384;38	Séquençage de Nouvelle Génération (NGS) des exomes par synthèse (Illumina) 5]					
3.2.4 [384;38	Séquençage de Nouvelle Génération (NGS) ciblé par synthèse (Illumina) 5]					

	3.2.	5 Séquençage Sanger	0
4	RÉ	SULTATS	2
	4.1	Etude 1 : A propos de 17 formes familiales de cancer du sein	2
	4.1.	1 Recherche de régions candidates par Génotypage tumoral	2
	4.1.	2 Recherche de variants candidats par Séquençage d'exome	5
	4.2 mutati	Etude 2 : A propos de 11 Formes familiales de cancer du sein associées à une ion constitutionnelle du gène <i>ATM</i>	8
5	DIS	CUSSION ET PERSPECTIVES 174	4
	5.1 prédis	L'apport des explorations tumorales dans l'identification de nouveaux gènes de position héréditaire au cancer	4
	5.2	L'implication du gène ATM dans la prédisposition héréditaire au cancer du sein 178	8
	5.3	L'hérédité manquante dans les formes familiales de cancer du sein18	1
6	RÉI	FÉRENCES BIBLIOGRAPHIQUES 182	3
7	UR	L	2
8	AN	NEXES	3

Table des figures

Figure 1: Classification hiérarchique supervisée selon la liste des « gènes intrinsèques » de Sorlie
Figure 2: Relation entre la pénétrance de la maladie, la fréquence allélique et les stratégies d'identification des facteurs génétiques de prédisposition au cancer du sein
Figure 3: Structures et protéines partenaires de BRCA1 et BRCA2
Figure 4: Motifs communs aux PI3KKs 55
Figure 5: Rôle de la protéine ATM dans la réparation des cassures double-brin de l'ADN 55
Figure 6: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995049
Figure 7: Profils génomiques obtenus par puces SNP pour les trois cancers du sein de la famille G1996019
Figure 8: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2001185
Figure 9: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2003221
Figure 10: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2007180
Figure 11: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009223
Figure 12: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995064 et absence de mise en évidence de région de LOH commune
Figure 13: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2007398 et absence de mise en évidence de région de LOH commune
Figure 14: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2010220 et absence de mise en évidence de région de LOH commune
Figure 15: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009373 (Témoin <i>BRCA2</i>) et absence de mise en évidence de région de LOH commune au locus du gène <i>BRCA2</i> (13q12-13)

Figure 16: Profils génomiques obtenus par puces SNP pour les trois cancers du sein (lié 2 à 2) de la famille G1995037 et mise en évidence de régions de LOH communes
Figure 17: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995042 et mise en évidence de régions de LOH communes
Figure 18: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995079 et mise en évidence de régions de LOH communes
Figure 19: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2002199 et mise en évidence de régions de LOH communes
Figure 20: Profils génomiques obtenus par puces SNP pour les trois cancers du sein de la famille G2006232 et mise en évidence de régions de LOH communes
Figure 21: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009159 et mise en évidence de régions de LOH communes
Figure 22: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2010325 et mise en évidence de régions de LOH communes 100
Figure 23: Exemple de visualisation de trois haplotypes conservés pour les trois tumeurs du sein (liées 2 à 2) de la famille G1995037 108
Figure 24: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G1995042 109
Figure 25: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G1995079 110
Figure 26: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G2002199
Figure 27: Exemple de visualisation d'un haplotype conservé pour les trois tumeurs du sein de la famille G2006232
Figure 28: Exemple de visualisation de deux haplotypes conservés pour les deux tumeurs du sein de la famille G2009159
Figure 29: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G2010325
Figure 30: Procédure de tri des variants dans le logiciel ERIS® d'IntegraGen 116

Liste des tableaux

Tableau 1: Classification des carcinomes infiltrants du sein selon l'OMS 2012 19
Tableau 2: Caractéristiques des 17 formes familiales de cancer du sein
Tableau 3: Caractéristiques des 45 régions de LOH identifiées 101
Tableau 3: Caractéristiques des 45 régions de LOH identifiées (suite) 102
Tableau 4: Caractéristiques des 7 haplotypes conservés détectés pour les trois cancers du sein analysés dans la famille G1995037 103
Tableau 5: Caractéristiques des 3 haplotypes conservés détectés pour les deux cancers du sein analysés dans la famille G1995042
Tableau 6: Caractéristiques des 7 haplotypes conservés détectés pour les deux cancers du sein analysés dans la famille G1995079 104
Tableau 7: Caractéristiques des 4 haplotypes conservés détectés pour les deux cancers du sein analysés dans la famille G2002199
Tableau 8: Caractéristiques des 19 haplotypes conservés détectés pour les trois cancers dusein analysés dans la famille G2006232105
Tableau 9: Caractéristiques des 22 haplotypes conservés détectés pour les deux cancers dusein analysés dans la famille G2009159106
Tableau 10: Caractéristiques des 12 haplotypes conservés détectés pour les deux cancers dusein analysés dans la famille G2010325107
Tableau 11: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G1995037_III:12
Tableau 12: Variants detectés par séquençage d'exome dans les haplotypes conservésidentifiés chez la patiente G1995042_II:14119
Tableau 13: Variants detectés par séquençage d'exome dans les haplotypes conservésidentifiés chez la patiente G1995079_III:1120
Tableau 14: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2002199_II:18

Tableau 15: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2006232_II:26	122
Tableau 16: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2009159_II:25	123
Tableau 16: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2009159_II:25 (suite)	124
Tableau 17: Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2010325_II:8	125

Liste des Annexes

Annexe 1: Arbre généalogique de la famille G1995049 (Témoin BRCA1)	
Annexe 2: Arbre généalogique de la famille G1996019	
Annexe 3: Arbre généalogique de la famille G2001185	
Annexe 4: Arbre généalogique de la famille G2003221 (Témoin BRCA1)	
Annexe 5: Arbre généalogique de la famille G2007180	
Annexe 6: Arbre généalogique de la famille G2009223	
Annexe 7: Arbre généalogique de la famille G1995064	220
Annexe 8: Arbre généalogique de la famille G2007398	221
Annexe 9: Arbre généalogique de la famille G2010220	222
Annexe 10: Arbre généalogique de la famille G2009373 (Témoin BRCA2)	223
Annexe 11: Arbre généalogique de la famille G1995037	
Annexe 12: Arbre généalogique de la famille G1995042	225
Annexe 13: Arbre généalogique de la famille G1995079	
Annexe 14: Arbre généalogique de la famille G2002199	227
Annexe 15: Arbre généalogique de la famille G2006232	
Annexe 16: Arbre généalogique de la famille G2009159	229
Annexe 17: Arbre généalogique de la famille G2010325	
Annexe 18: Caractéristiques des couples d'amorces utilisés pour l'amplification des gène <i>ATM</i>	exons du 231
Annexe 19: Composition du mix réactionnel des PCR	
Annexe 20: Programme des PCR du gène ATM	
Annexe 21: Composition du mix des réactions de séquence	

Liste des abréviations

- ADN : Acide désoxyribonucléique
- BAC : Bacterial Artificial Chromosome
- BAF : B Allele Frequency
- bam : Version binarisée du format sam
- BRCA1 : BReast Cancer 1
- BRCA2 : BReast Cancer 2
- BRCT : BRCA1 C-Terminus
- CDB : Cassure double-brin
- CGH : Comparative Genomic Hybridization
- CRT : Cyclic reversible termination
- ERa : Estrogen Receptor alpha
- GAP : Genome Alteration Print
- HMZ : Homozygote
- HTZ : Hétérozygote
- IBD : Identity By Descent
- INDEL : Insertion-délétion
- kb : kilobase
- LOD score : Logarithm of odds score
- LOH : Loss of Heterozygosity
- LRR : Log R ratio
- NGS : Next Generation Sequencing
- PARP1 : Poly(ADP-ribose) polymérase 1
- PCR : Polymerase Chain Reaction
- pb : paire de bases
- RA : Récepteurs aux Androgènes
- RE : Récepteurs aux Estrogènes
- **RH** : Récepteurs Hormonaux
- RP : Récepteurs à la Progestérone

RR : Risque Relatif sam : Sequence Alignment/Map SNP : Single Nucleotide Polymorphism SNV : Single Nucleotide Variant vcf : Variant call format

Résumé

5 à 10% des cancers du sein sont héréditaires mais parmi ceux-ci seulement la moitié est expliquée par une altération constitutionnelle d'un gène de prédisposition connu tels que les gènes BRCA1 et BRCA2. L'importante hétérogénéité génétique qui caractérise les familles BRCAx rend difficile la réalisation d'études familiales groupées et ne permet pas l'identification de nouveaux gènes de prédisposition au cancer du sein selon les méthodes classiques de liaison génétique ou d'association. Les techniques de séquençage de nouvelle génération (NGS) à l'échelle de l'exome ou du génome entier, autorisent en revanche l'étude de familles individuelles à la recherche de mutations constitutionnelles privées mais le nombre considérable de variants génétiques identifiés impose leur tri sur des critères de pathogénicité ou de récurrence. Un autre critère de tri peut être représenté par l'identification de régions candidates définies en fonction de réarrangements génomiques tumoraux communs à plusieurs tumeurs au sein d'une même famille. Le génotypage tumoral par puces SNP (pour single nucleotide polymorphism) permet en effet la détection d'haplotypes conservés dans des régions récurrentes de LOH (pour loss of heterozygosity) communes à plusieurs tumeurs familiales et donc l'identification de régions candidates suspectes d'abriter des mutations germinales dans des gènes de prédisposition au cancer.

La combinaison de ces deux approches, génotypage tumoral puis NGS, a été appliquée à une série de 17 familles avec agrégation de cancers du sein pour lesquelles au moins deux échantillons tumoraux étaient disponibles. Aucun nouveau gène de prédisposition au cancer du sein n'a été identifié mais une mutation délétère constitutionnelle du gène *ATM* a ainsi été retrouvée, associée à une perte de l'allèle sauvage dans les 2 tumeurs d'une famille BRCAx.

L'analyse de 17 tumeurs du sein supplémentaires provenant de 10 familles avec agrégation de cancers du sein et mutation constitutionnelle du gène *ATM* identifiée chez le cas index, a révélé que l'allèle sauvage d'*ATM* était fréquemment perdu dans ces tumeurs (>80% contre 20% attendu en situation sporadique ; p<0.001). Ce résultat plaide fortement en faveur de l'implication d'*ATM* dans la carcinogénèse de ces cancers du sein tel un gène suppresseur de tumeur et suggère que les mutations constitutionnelles d'*ATM* sont impliquées dans des formes familiales de cancer du sein.

Mots clés: Cancer du sein familial, ATM, puces SNP, NGS, Risque de cancer du sein

Abstract

Hereditary breast cancers (BCs) account for 5-10% of all diagnosed BCs, yet only 50% of such tumors arise in the context of a germline mutation in known tumor suppressor genes such as *BRCA1* or *BRCA2*. The vast genetic heterogeneity which characterizes BRCAx families makes grouped studies impossible to perform. Next generation sequencing (NGS) techniques, however, allow individual families to be studied in order to identify private mutations. Single nucleotide polymorphism (SNP) arrays allow the detection of conserved haplotypes within recurrent regions of loss of heterozygosity, common to several familial tumors, therefore identifying genomic loci likely to harbor a germline mutation in cancer predisposition genes.

The combination of both exome sequencing and SNP arrays for a series of 17 familial BC did not allow the identification of a novel BC predisposition gene, but revealed a germline *ATM* mutation associated with a loss of the wild-type allele in a BRCAx family.

The analysis of 17 additional breast tumors from ten BC families in which a germline ATM mutation had been identified revealed a high frequency of wild-type allele loss in these tumors (>80% compared to the 20% expected in sporadic BC; p <0.001). This result argues strongly in favor of the involvement of ATM in the carcinogenesis of these tumors as a tumor suppressor gene and suggests that germline ATM mutations are involved in a subset of familial BC.

Keywords: Familial breast cancer, ATM, SNP array, NGS, Breast cancer risk

1 INTRODUCTION

1.1 Epidémiologie du cancer du sein

Le cancer du sein se situe au premier rang des cancers incidents chez la femme, nettement devant le cancer colorectal et le cancer du poumon, et particulièrement dans les pays industrialisés avec plus d'un million de nouveaux cas par an dans le monde [1]. De nos jours, on considère qu'une femme sur huit développera un cancer du sein au cours de sa vie. D'après l'Institut National du Cancer (INCa), 54062 nouveaux cas de cancer du sein auraient été diagnostiqués en 2015 en France métropolitaine [2]. Comparativement, le cancer du sein chez l'homme est extrêmement rare et concerne moins de 1% de tous les cancers du sein. Bien qu'il soit de bon pronostic avec une survie nette standardisée à 5 ans de 87%, le cancer du sein constitue la première cause de décès par cancer chez la femme avec plus de 11000 décès (11913 estimés en 2015) recensés par an en France.

L'incidence du cancer du sein, partant de 21000 nouveaux cas en 1980, a plus que doublé au cours de ces trente dernières années [3;4], probablement à cause de l'augmentation de l'espérance de vie, d'un déploiement du dépistage, mais également de la mise en place des traitements hormonaux de la ménopause combinant œstrogènes et progestérone. Depuis 2005, son incidence semble diminuer. En effet, le taux d'incidence standardisé a augmenté de + 1,4 % par an en moyenne entre 1980 et 2012, mais on observe une diminution de - 1,5 % par an si on se focalise sur la période située entre 2005 et 2012 (56,3 cas pour 100 000 personnes-années en 1980, 97,8 en 2005 et 88,0 en 2012) [2]. Le taux d'incidence a de plus baissé rapidement sur une courte période entre 2005 et 2007-2008. Cette diminution décrite dans plusieurs pays occidentaux, concernait principalement les femmes de 50 à 69 ans [5]. Parmi les facteurs pouvant expliquer cette baisse, la saturation du dépistage et la diminution massive de la prescription et de l'utilisation des traitements hormonaux de la ménopause ont été avancées [6;7].

1.2 Étiologies du cancer du sein

Le cancer du sein est une maladie multifactorielle. Cela signifie que plusieurs facteurs influent sur le risque de sa survenue. On parle de facteurs de risque. De nombreux facteurs de risque du cancer du sein sont connus. Certains ont fait l'objet en 2008, à la demande de l'AFSSET (Agence Française de Sécurité Sanitaire de l'Environnement et du Travail), d'un rapport d'expertise réalisé par l'Inserm et intitulé « Cancer et Environnement » [8]. Ils sont présentés ci-après même s'il existe encore aujourd'hui des incertitudes quant à l'implication et au poids de plusieurs de ces facteurs. Une personne présentant un ou plusieurs facteurs de risque peut effectivement ne jamais développer de cancer et inversement, il est possible qu'une personne n'ayant aucun facteur de risque soit atteinte de ce cancer.

1.2.1 L'âge

Comme c'est le cas pour beaucoup de cancers de l'adulte, l'âge constitue un facteur de risque important dans la survenue d'un cancer du sein [9]. La maladie est effectivement rare chez les femmes de moins de 30 ans. En revanche, le risque augmente nettement entre 50 et 75 ans où près de deux tiers des cancers du sein sont diagnostiqués, avec un pic de fréquence autour de 60 ans [10;11] et un âge médian au diagnostic estimé en France à 63 ans [2]. C'est pour cette raison que toutes les femmes de cette tranche d'âge sont invitées, dans le cadre du dépistage organisé du cancer du sein, à bénéficier d'une mammographie tous les 2 ans, prise en charge par l'Assurance maladie.

1.2.2 Le mode de vie

Certains éléments du mode de vie sont associés à une modification, positive ou négative, du risque de cancer du sein.

Une activité physique régulière apparaît ainsi comme un facteur protecteur [12;13] alors que la consommation d'alcool, même modérée et équivalent à la prise d'un verre par jour, augmente le risque de cancer du sein, quel que soit le type de boisson [14-17].

De même, la consommation excessive d'acides gras saturés et la faible consommation de fruits et de légumes constitueraient des facteurs aggravants sans que les données épidémiologiques soient significatives dans l'ensemble des études [18].

Concernant l'impact du tabagisme sur les risques de cancer du sein, les données sont plus controversées et largement détaillées dans le rapport d'expertise Inserm pour l'AFSSET [8]. Le tabac peut être considéré comme un anti-estrogène, accélérant le catabolisme de l'œstradiol [19]. Ainsi, une étude a montré une diminution significative du risque de cancer du sein dans le groupe fumeur d'une population de femmes porteuses de mutations *BRCA1* et *BRCA2* par rapport au groupe non-fumeur [20]. Une analyse plus récente a cependant indiqué que le risque de cancer du sein était légèrement augmenté chez les femmes fumant avant la ménopause et plus particulièrement avant la première grossesse [21], laissant penser que l'effet du tabac, à la fois cancérogène et anti-estrogène, pourrait varier au cours de la vie et selon le début de l'exposition.

D'autres facteurs de risque liés à notre environnement ont été identifiés récemment, tels que certains produits chimiques ou xénohormones qui pourraient interférer avec les voies de signalisation des hormones endocriniennes et augmenter le risque de cancer du sein [22].

1.2.3 Les maladies bénignes du sein

Les maladies bénignes du sein constituent un facteur de risque de cancer du sein. Elles sont histologiquement divisées en deux groupes : les lésions prolifératives et les lésions non prolifératives avec ou sans atypie. Les lésions non prolifératives ne sont généralement pas associées à un risque accru de cancer du sein ou, si elles le sont, le risque est très faible. Les lésions prolifératives sans atypie multiplient le risque par deux, tandis que les lésions hyperplasiques avec atypies augmentent ce risque d'au moins quatre fois [23;24].

1.2.4 La densité mammaire

La densité mammaire évaluée par mammographie donne une estimation de la composition relative du tissu mammaire en tissu stromal, épithélial et adipeux. Pour des femmes du même

âge, une importante densité mammaire due à une forte proportion de stroma et d'épithélium, est associée à un risque relatif de quatre à six [25;26].

1.2.5 L'exposition aux radiations ionisantes

Les radiations ionisantes augmentent le risque de cancer car elles induisent des dommages à l'ADN et à ses constituants. Un suivi intensif de plusieurs groupes de population a montré que le sein était l'un des organes les plus sensibles aux effets des radiations [23].

Le risque de cancer du sein est d'autant plus augmenté que l'âge de la femme est précoce au moment de l'exposition. Un excès de cancers du sein a ainsi été observé chez des femmes exposées aux radiations de bombes atomiques durant la seconde guerre mondiale alors qu'elles étaient âgées entre 10 et 19 ans [27]. Il a également été montré que l'effet des radiations ionisantes, chez les femmes exposées avant l'âge de 40 ans, était associé à un risque de cancer du sein multiplié par trois, pour une exposition évaluée à 1 Gy [28], mais ce risque semble diminuer fortement avec l'âge à l'exposition aux radiations ionisantes [8]. Le risque de cancer du sein est par ailleurs similaire pour une exposition unique ou pour des expositions multiples à intensité totale égale [29].

Deux études américaines ont également montré que les examens radiographiques thoraciques répétés chez des jeunes filles suivies pour une scoliose depuis l'âge de 10-12 ans jusqu'à l'âge de 40-50 ans en moyenne, augmentaient de manière significative leur risque de cancer du sein pour des doses cumulées aussi faibles que 130 mGy [30;31].

Plusieurs études ont essayé de comparer les bénéfices et les risques du dépistage par mammographies en fonction de l'âge de début de dépistage. D'une manière générale, les résultats quantitatifs dépendent des coefficients et des modèles de risque utilisés ainsi que des hypothèses concernant l'efficacité du dépistage et les variations d'agressivité des tumeurs en fonction de l'âge. Cependant, la plupart des auteurs concluent que pour un âge de début de dépistage entre 40 et 50 ans, le bénéfice est nettement plus important que le risque [8].

Aussi, des cancers du sein peuvent survenir dans les années suivant un traitement par irradiation, notamment dans le cadre d'une maladie de Hodgkin [32;33].

1.2.6 L'exposition aux hormones endogènes et exogènes

Les études cliniques, animales et épidémiologiques, ont clairement démontré le caractère hormono-dépendant du cancer du sein et plusieurs facteurs relatifs au statut hormonal ont été trouvés associés au risque de cancer du sein [34].

Ainsi, le risque de cancer du sein est accru par un âge précoce aux premières règles, la nulliparité, une première grossesse à terme et une ménopause tardives [35]. L'exposition aux hormones endogènes [36-38], de même que l'utilisation d'hormones exogènes sous forme de contraceptifs oraux [39], de traitements hormonaux de la ménopause [40;41] ou de traitements de la stérilité [42], sont associées à une augmentation de risque de cancer du sein. Inversement, le risque de cancer du sein est diminué par un allaitement prolongé [43]. Une parité précoce aurait également un effet protecteur à long terme [27] alors que la grossesse elle-même, serait associée à un risque transitoire [44]. Les grossesses non menées à terme (fausses couches, interruptions volontaires ou thérapeutiques de grossesse) ne semblent pas influencer le risque de cancer du sein [45].

L'obésité est elle-même associée à un profil hormonal soupçonné de favoriser le développement du cancer du sein. Elle augmenterait d'environ 50% le risque de cancer du sein chez les femmes ménopausées, probablement en raison de l'augmentation des concentrations sériques d'œstradiol libre [23]. En effet, l'excès de tissu adipeux entraîne l'augmentation de la production et du temps d'exposition aux hormones stéroïdiennes [46]. Après la ménopause, l'aromatisation des androgènes dans le tissu adipeux est l'une des plus considérables sources d'œstrogènes circulants. En revanche, parce qu'elle donne souvent lieu à des cycles menstruels anovulatoires, l'obésité n'augmente pas le risque chez les femmes avant la ménopause. Elle serait même associée à un risque réduit chez ces femmes dans les pays économiquement développés.

Le rôle des œstrogènes dans la cancérogénèse mammaire est de plus mis en évidence par l'efficacité de l'ovariectomie et de l'hormonothérapie par tamoxifène et inhibiteurs d'aromatase, dans la prévention et le traitement des cancers du sein. Les mécanismes oncogéniques des œstrogènes ne sont pas entièrement élucidés. Ils seraient cependant liés aux capacités de stimulation de la prolifération cellulaire et d'inhibition de l'apoptose ainsi qu'aux

dommages à l'ADN causés par les métabolites réactifs issus du métabolisme des œstrogènes [47].

1.2.7 Les antécédents de cancer du sein

Les antécédents personnels et familiaux de cancer du sein constituent un risque majeur de cancer du sein. En effet, une femme ayant développé un cancer du sein présente un risque deux à quatre fois plus élevé d'en développer un nouveau qu'une femme indemne de cancer au même âge. Aussi, une femme ayant un, deux ou trois parents au premier degré (mère, sœur ou fille) atteint(s) d'un cancer du sein multiplie respectivement par deux, trois ou quatre son risque de cancer du sein [48]. Le risque individuel de cancer du sein augmente ainsi avec le nombre d'apparentés atteints et l'âge précoce au diagnostic [48;49].

Près de 20 à 30 % des cancers du sein se manifestent chez des femmes ayant des antécédents familiaux de cancers du sein [50;51]. Le risque augmenté associé aux antécédents familiaux évoque une prédisposition familiale. Cependant, il n'est pas simple de savoir si cette disposition familiale est due au hasard, à un mode de vie que des membres de la famille ont en commun, à un facteur héréditaire (ou mutation génétique) constitutionnel transmis des parents aux enfants, ou s'il s'agit d'une association de ces différents éléments. Il est toutefois admis que 5 à 10% des cancers du sein sont héréditaires [52]. Deux gènes majeurs de susceptibilité au cancer du sein ont été identifiés dans les années 1990 : *BRCA1* (pour *Breast Cancer 1*) et *BRCA2* (pour *Breast Cancer 2*). Leur implication dans la prédisposition héréditaire au cancer du sein ainsi que celle d'autres gènes découverts depuis, seront présentées dans le chapitre 1.4 de l'introduction.

1.3 Classification des cancers du sein

Le cancer du sein est une maladie très hétérogène dans le sens où les tumeurs mammaires présentent une grande variété phénotypique d'un cancer à l'autre mais aussi au sein d'une même tumeur, en fonction des structures mammaires dans lesquelles elles prennent naissance, de leur stade de progression, et des marqueurs biologiques exprimés dans chaque cellule tumorale. Les classifications des cancers du sein ont évolué au cours du temps et des méthodes d'analyse.

1.3.1 Classification histologique, phénotypique et pronostique

Classification histologique

Les premières classifications des cancers du sein se sont basées sur des observations histologiques et morphologiques. La classification actuellement utilisée est celle révisée en 2012 par l'Organisation Mondiale de la Santé (OMS) [53].

Les tumeurs épithéliales malignes ou carcinomes représentent la grande majorité des tumeurs malignes du sein. Les tumeurs malignes non épithéliales (sarcomes, lymphomes malins non hodgkiniens primitifs) sont en revanche beaucoup plus rares (moins de 1% des cancers du sein)

D'un point de vue histologique, les tumeurs peuvent tout d'abord être classées en fonction de l'envahissement de la glande mammaire par les cellules néoplasiques. On distingue ainsi (i) les cancers in situ, se définissant comme une prolifération néoplasique épithéliale restreinte au compartiment de la glande mammaire délimité par la membrane basale, sans risque théorique d'envahissement ganglionnaire, et (ii) les cancers infiltrants, se caractérisant par l'envahissement du tissu conjonctif par les cellules néoplasiques au-delà de la membrane basale.

La classification histologique des carcinomes infiltrants du sein revue récemment par l'OMS, définit une vingtaine d'entités différentes (Tableau 1). Les deux principaux types sont le type canalaire (appelé désormais le type non spécifique ou TNS) et le type lobulaire représentant respectivement 40 à 75% et 5 à 15% des cas [53].

Les *tumeurs de type lobulaire* se caractérisent par la perte de l'E-cadhérine, protéine impliquée dans la cohésion des cellules épithéliales entre-elles. Sur le plan morphologique, ceci se traduit par de petites cellules tumorales, généralement régulières et peu mitotiques, comportant une inclusion de mucus dans le cytoplasme. Ces cellules non cohésives, disposées

Tableau 1. Classification des carcinomes infiltrants du sein selon l'OMS 2012

- Carcinome canalaire infiltrant ou de type non spécifique (TNS)
- Carcinome lobulaire infiltrant
- Carcinome tubuleux
- Carcinome cribriforme
- Carcinome mucineux
- Carcinome médullaire
- Carcinome à différentiation apocrine
- Carcinome à cellules en bague à chaton
- Carcinome micropapillaire infiltrant
- Carcinome métaplasique
- Carcinome neuroendocrine
- Carcinome sécrétant
- Carcinome papillaire infiltrant
- Carcinome à cellules acineuses
- Carcinome mucoépidermoïde
- Carcinome polymorphe
- Carcinome oncocytique
- Carcinome à cellules riches en lipides
- Carcinome à cellules claires (riches en glycogène)
- Carcinome sébacé
- Tumeurs des glandes salivaires et des glandes annexielles
- Carcinome adénoïde kystique
- Carcinome inflammatoire

le plus souvent en « file indienne », envahissent le stroma mammaire en respectant les structures préexistantes. Le pronostic de ces cancers longtemps considéré comme meilleur, rejoint finalement celui des canalaires infiltrants [53].

Les *tumeurs de type canalaire ou TNS* reproduisent l'architecture des acini mammaires. Ils se définissent sur le plan morphologique par l'absence de signe histologique spécifique d'un autre type sur plus de 50% de leur surface. Les cellules tumorales sont cohésives, le stroma est d'abondance et de nature variées (myofibroblates ou fibreux, infiltrat inflammatoire), la densité vasculaire est hétérogène d'une tumeur à l'autre et l'association à une composante carcinomateuse in situ est observée dans 40 à 50% des cas [53].

Le *carcinome médullaire* est une autre de forme de cancer infiltrant, qui représente environ 1 à 5% des cancers du sein et dont la définition histologique répond aux cinq critères définis par Ridolfi et al. en 1977 : plus de 75% de circonscription de la tumeur, architecture syncitiale dans plus de 75% de la tumeur, stroma lymphoplasmocytaire modéré à marqué, atypies nucléaires modérées à marquées, absence de différenciation glandulaire [54]. Il touche souvent des femmes âgées de moins de 50 ans et est souvent retrouvé chez les patientes ayant une mutation du gène *BRCA1*. Ce type de cancer est plutôt associé à un bon pronostic [53].

D'autres formes de carcinome infiltrant existent, dites de bon pronostic comme le *carcinome tubuleux*, composé essentiellement de petits tubes à une seule couche de cellules tumorales régulières ; le *carcinome mucineux* qui est une forme rare de la femme âgée, comportant de larges flaques de mucus extracellulaire au sein desquelles flottent des ilots de cellules malignes ; le *carcinome adénoïde cystique ou cylindrome*, comportant comme au niveau des glandes salivaires, une prolifération biphasique de petites cellules basaloïdes et un contingent de cellules épithéliales, l'ensemble s'agençant sur un mode cribriforme, tubulaire, trabéculaire ou massif ; le *carcinome cribriforme* infiltrant constitué en majorité de travées infiltrantes d'architecture cribriforme avec atypies légères à modérées [53;55].

Certains types de cancer sont au contraire connus pour leur mauvais pronostic comme le *carcinome inflammatoire*, défini sur des critères cliniques (sein rouge, chaud, aspect de peau d'orange, évolutivité rapide de la tumeur) et/ou pathologiques comme la présence d'emboles dans les lymphatiques du derme superficiel [53;55].

Le grade histopronostique

L'analyse histologique d'une tumeur fournit des informations sur le pronostic associé. La classification originale de Scarff, Bloom et Richardson [56] est basée sur trois critères histologiques différents : (i) le degré de différenciation architecturale (la formation de tubes), (ii) l'importance du pléomorphisme nucléaire et (iii) le nombre de mitoses par champ. A chacun de ces critères est attribué un score de 1 à 3 dont la somme constitue le score histopronostique. Le grade I correspond ainsi à un score inférieur à 6, le grade II à un score de 6 à 7 et le grade III, à un score de 8 à 9. Plus le grade de la tumeur est élevé, moins le pronostic associé est bon.

Cette classification a été modifiée ensuite par Elston et Ellis [57] et adoptée de façon plus générale dans le cadre des critères de Nothingham [58;59]. En effet, la taille de la tumeur et le statut des ganglions axillaires (score de 1 : absence d'envahissement ganglionnaire ; score de 2 : atteinte de un à trois ganglions ; score de 3 : atteinte de plus de trois ganglions) sont également pris en compte pour l'évaluation du pronostic. La somme du score ganglionnaire, du grade histopronostique et de la taille de la tumeur (en cm) pondérée de 20%, représente l'index pronostique de Nottingham ou NPI (pour Nottingham Prognostic Index).

Le statut des récepteurs hormonaux

La détermination du statut des récepteurs hormonaux, œstrogènes (RE) et progestérone (RP), participe également à la classification des carcinomes mammaires infiltrants ainsi qu'à l'évaluation du pronostic et à la prise en charge des patientes. Le niveau d'expression des récepteurs aux œstrogènes est inversement corrélé au grade histopronostique et à l'expression de ERBB2 [60]. 70 à 80% des carcinomes mammaires expriment les récepteurs hormonaux, permettant la mise en place d'une hormonothérapie [53]. La valeur pronostique de l'expression du récepteur à la progestérone est également forte. Elle est indépendante du grade et de la taille de la tumeur en particulier chez les patientes sans envahissement ganglionnaire. De plus, la survie à 5 ans est proportionnelle au niveau d'expression du récepteur à la progestérone [61].

Le statut de l'Oncogène HER2

L'amplification de l'oncogène HER2 (codé par le gène *ERBB2*) a été décrite à la fin des années 80 [62]. La détermination du statut de l'oncogène HER2 est une information importante pour la prise en charge médicale des patientes depuis la découverte d'un anticorps, le trastuzumab ou Herceptin®, ciblant de manière spécifique les cellules sur-exprimant cet oncogène [63] dont la commercialisation a été autorisée en Europe en 2000. En effet, des essais thérapeutiques ont démontré une réduction importante des rechutes et de métastases (environ 50%) chez les patientes bénéficiant d'un traitement par Herceptin® en plus de la chimiothérapie [64;65].

La diversité des types histologiques se traduit ainsi par une hétérogénéité de pronostic. Le type histologique d'un cancer du sein n'est cependant qu'un paramètre pronostique parmi d'autres comme la taille tumorale, le grade et le statut ganglionnaire. Nous allons voir également qu'une donnée plus récente telle que le type moléculaire participe à la classification des cancers mammaires ainsi qu'à leur évaluation pronostique.

1.3.2 Classification moléculaire des cancers du sein

L'hétérogénéité des cancers du sein a rendu nécessaire l'établissement de classifications permettant la création de groupes homogènes de tumeurs, ceci dans une perspective pronostique et diagnostique. Les nouvelles classifications, dites moléculaires, ont été développées grâce à l'apparition d'outils regroupés sous le terme générique de biopuces. Ces techniques permettent l'analyse du génome à grande échelle ainsi que l'analyse du transcriptome par appréciation de l'expression des gènes. Elles ont pour but commun l'identification de facteurs pronostiques et prédictifs utilisables par la suite dans la prise en charge du malade.

L'hypothèse de l'intérêt de ces analyses moléculaires formulée par l'équipe de Charles M. Perou est que la diversité phénotypique des cancers du sein peut être accompagnée d'une diversité au niveau de l'expression des gènes, étudiée par micro-array [66]. C'est à cette équipe que l'on doit une première classification moléculaire des cancers du sein en quatre sous-types sur des données transcriptomiques de 65 tumeurs, affinée à l'aide d'études successives en 2001, 2003 et 2004 sur un plus grand nombre de prélèvements, et définissant les cinq sous-types suivants : Basal-like, ERBB2+, Normal Breast-like, Luminal B et Luminal

A (Figure 1) [67-69]. Cette classification a rapidement été associée à une valeur pronostique et à différents taux de mutation de *TP53* et autres gènes associés au cancer du sein [67;70]. Ces sous-types appartiennent à deux grands groupes : un groupe caractérisé par une expression plus ou moins forte du récepteur aux œstrogènes codé par *ESR1 (estrogen receptor 1)*, tumeurs dites « RE positives » et l'autre par une expression faible ou nulle de celui-ci, tumeurs dites « RE négatives ».

1.3.2.1 Le groupe des tumeurs « RE positives »

Le groupe des tumeurs « RE positives » rassemble les tumeurs luminales A et luminales B.

Les tumeurs luminales A sont des cancers de bas grade présentant une faible prolifération et une forte expression de *ESR1*, *GATA3* (*GATA binding protein 3*), *HNF3α* (*Hepatocyte Nuclear Factor 3α*), *XBP1* (*X-box binding protein 1*) et *LIV-1* (*oestrogen regulated protein LIV1*) [69]. Elles présentent un faible taux de mutation de *TP53* (< 20%) [67;69]. D'un point de vue fonctionnel, ces tumeurs expriment principalement des gènes impliqués dans le métabolisme des acides gras et dans la voie de signalisation des récepteurs aux œstrogènes [71]. Le nombre de gènes fréquemment mutés s'avère être le plus important dans ce sous-groupe avec les gènes PIK3CA (45%), *GATA3* (14%), *MAP3K1* (*MAP kinase kinase kinase 3*) (13%), *TP53* (12%), *CDH1* (9%), *MLL3* (*mixed-lineage leukemia 3*) (8%) et *MAP2K4* (7%) [70]. Les mutations inactivatrices de *MAP3K1* et *MAP2K4* sont mutuellement exclusives. Ces deux gènes sont des acteurs de la voie de stress p38-JNK1 (*Jun N-terminal Kinase (JNK) 1*).

Les tumeurs luminales B sont de plus haut grade que les tumeurs luminales A, avec un index de prolifération plus important, et elles expriment en moindre quantité les récepteurs hormonaux RE et RP. Elles se caractérisent ainsi par une moindre expression des gènes exprimés par les tumeurs du groupe luminal A et une forte expression des gènes impliqués dans le cycle cellulaire tels que les gènes *CCNE1* et *CCND1* (Cycline E1 et D1), mais également les gènes *GGH* (*YGlutamyl Hydrolase Precursor*), *NSEP1* (*Nuclease Sensitive Element Binding Protein 1*), *LAPTM4β* (*Lysosomal protein transmembrane 4 β*) et *MYB* (*Myeloblastosis viral oncogene homolog*) [69;72]. Ces tumeurs présentent des mutations des gènes *PIK3CA* et *TP53* à une fréquence de 29%, des mutations du gène *GATA3* à une



Figure 1: Classification hiérarchique supervisée selon la liste des « gènes intrinsèques » de Sorlie (d'après Sorlie et al, Proc Natl Acad 2003 [68]).

(A) Représentation du cluster entier des 534 gènes et 122 échantillons (115 prélèvements tumoraux et 7 tissus non malins). (B) Dendrogramme montrant la classification des tumeurs du sein des 5 sous-groupes dont les branches sont colorées en fonction du sous-type moléculaire : « luminal A » en violet, « luminal B » en bleu, « ERBB2+ » en rose, « basallike » en orange et « Normal Breast-like » en vert. fréquence de 15%, et une fréquence faible de mutations des gènes *MAP3K1* (5%), *CDH1* (5%) et *MLL3* (6%) [70].

1.3.2.2 Le groupe des tumeurs « RE négatives »

Le groupe des tumeurs « RE négatives » comprend les tumeurs dites basales, les tumeurs ERBB2 positives (par amplification du gène *ERBB2*) et les tumeurs dites « normales like ».

Les tumeurs « basal-like » sont des tumeurs de haut grade histologique, associant une forte anisocaryose, un faible degré de différenciation et un index mitotique élevé [72;73]. Elles représentent environ 15% des carcinomes canalaires infiltrants. L'appellation « basal-like » provient du fait que ces tumeurs expriment de nombreux marqueurs également exprimés par les cellules basales de la glande mammaire normale. Ce sous-type correspond le plus souvent aux tumeurs dites « triple-négatives» du fait de l'absence d'expression des récepteurs hormonaux (RE et RP) et de l'absence de surexpression de HER2. Il se caractérise par l'absence d'expression des gènes associés au groupe « luminal » et par une forte expression des gènes d'origine myoépithéliale, notamment KRT5 (cytokératine 5), KRT6B (cytokératine 6B), KRT14 (cytokératine 14), KRT17 (cytokératine 17), LAMC2 (laminine Y2), CAV1 (cavéoline 1), CAV2 (cavéoline 2), FABP7 (Fatty Acid Binding Protein 7), et EGFR (Epidermal Growth Factor Receptor) [71;72;74]. D'un point de vue fonctionnel, ces tumeurs expriment principalement des gènes impliqués dans le contrôle du cycle cellulaire, de la prolifération et de la réponse immunitaire [71;72]. Ce groupe présente à la fois le plus haut taux de mutations du gène TP53 (80%) et le plus bas taux de mutations du gène PIK3CA (9%). A noter que la fréquence du taux de mutations du gène MLL3 est de 5%, soit un taux semblable aux autres groupes [70]. Ce gène code pour une protéine appartenant au complexe ASCOM (ASC-2- and Mll3-containing complex) qui possède une activité histone méthyltransférase et il est décrit comme potentiel gène suppresseur de tumeur en raison de la fréquence de ses mutations dans divers types tumoraux et l'apparition de tumeurs lors de son inactivation chez des souris [75].

Les tumeurs ERBB2+ sont caractérisées par une surexpression du gène *ERBB2* (*Erythroblastic Leukemia Viral Oncogene Homolog 2*) et des autres gènes situés dans l'amplicon 17q22.24 comme *MED24* (*mediator complex subunit*) et *GRB7* (*growth factor*

receptor-bound protein 7) [69]. Ces tumeurs n'expriment pas les récepteurs hormonaux RE et RP. Elles présentent en revanche un taux élevé de mutations de *TP53* et *PIK3CA* s'élevant respectivement à 72% et 39% [70].

Les tumeurs « normal-like », sont caractérisées par une forte expression de gènes exprimés dans le tissu adipeux et les cellules non épithéliales.

Les cinq groupes moléculaires identifiés ont été corrélés à des différences de survie et de réponse au traitement. Les analyses des courbes de survie de Kaplan-Meier ont ainsi permis d'associer les sous-types basal-like et ERBB2+ au plus mauvais pronostic. Le sous-type luminal B présente également un moins bon pronostic que le sous-type luminal A [68]. La similarité du profil d'expression de certains gènes entre les sous-types luminal B , basal-like et ERBB2+ suggère que l'expression de ces gènes serait associée à un mauvais pronostic [69]. Ainsi, cette classification des cancers du sein en cinq sous-groupes apporte des valeurs pronostiques d'estimation du risque de rechute ou du bénéfice d'un traitement spécifique [76;77].

Suite à l'identification initiale de cinq sous-types moléculaires, une sous-classification a été proposée. Pour exemple, une analyse détaillée des gènes exprimés dans les tumeurs RE négatives montre que les tumeurs basales sont un groupe hétérogène de tumeurs pouvant être subdivisé en au moins trois sous-groupes: (i) le groupe « molecular apocrine » [78], qui partage les caractéristiques du sous-groupe « ERBB2+ » et semble présenter une activation de la voie de signalisation du récepteur aux androgènes (RA) ; (ii) le sous-type « interferon » [72], défini par une forte expression de STAT1 (*Signal transducer and Activator of Transcription 1*) et des gènes de l'immunité codant pour les interférons; (iii) le sous-groupe « claudin-low » qui inclut les tumeurs ayant un phénotype « *stem cell-like* » [79].

De plus, la classification non supervisée de Farmer et al, identifie trois groupes de tumeurs distincts [78]. Le groupe des tumeurs RE+ RA+ RP+ qui rassemble les tumeurs luminales, le groupe RE- RA- RP- qui contient les tumeurs basales et enfin le groupe RE- RA+, appelé moléculaire apocrine car il rassemble des tumeurs provenant de tissus engagés dans une différenciation apocrine. Ces dernières tumeurs sont en fait confondues avec les tumeurs ERBB2+ car elles présentent fréquemment l'amplicon 17q.

1.3.3 Caractéristiques des cancers du sein survenant dans le cadre d'une prédisposition héréditaire connue

1.3.3.1 Caractéristiques des cancers du sein liées à une mutation constitutionnelle de *BRCA1*

L'étude morphologique d'une grande série de tumeurs *BRCA1* réalisée par le Breast Cancer Linkage consortium a montré que la majorité de ces tumeurs (75%) est de type canalaire infiltrant [80]. Les carcinomes médullaires sont également particulièrement fréquents et représentent environ 13% des cas [81-83]. La présence d'une composante in situ associée au carcinome infiltrant est retrouvée dans 30 à 40% des cas [80;84]. Les tumeurs *BRCA1* sont fréquemment associées à un important infiltrat lymphocytaire et sont généralement de haut grade du fait d'une forte prolifération se traduisant par un index mitotique élevé et d'une faible différenciation [85-87]. En effet, environ 25% des tumeurs sont de grade II et 66% sont de grade III [80].

Sur le plan phénotypique, une des caractéristiques majeures des tumeurs *BRCA1* est l'absence d'expression des récepteurs hormonaux (œstrogènes et progestérone) et l'absence d'amplification de l'oncogène HER2, ce phénotype étant communément appelé « triple négatif » [88]. L'amplification de HER2 est effectivement rare dans les tumeurs *BRCA1*, observée dans 0 à 15% des cas selon les études [89;90], de même pour l'expression du récepteur aux œstrogènes et à la progestérone [86;89;91]. Les tumeurs *BRCA1* se caractérisent également par la forte expression du marqueur nucléaire Ki67 et la sous-expression de la cycline D1, illustrant leur forte prolifération [92;93].

1.3.3.1.1 Profil transcriptomique des tumeurs BRCA1

D'un point de vue transcriptomique, Hedenfalk et al ont montré en 2001 que les tumeurs *BRCA1*, *BRCA2* et sporadiques présentaient des signatures distinctes [94]. Cependant, suite à la mise en évidence des cinq sous-types tumoraux (normal-like, HER2, basal-like, luminal A et luminal B), il est rapidement apparu que la majorité des tumeurs *BRCA1* partageaient les caractéristiques phénotypiques des tumeurs de type basal-like, dont fait partie le phénotype « triple négatif ». En 2002, van't Veer et al, ont identifié une signature moléculaire permettant

d'identifier les tumeurs *BRCA1* à partir de données de transcriptome [95]. En réanalysant leurs données, Sorlie et al, ont montré que cette signature était en fait une signature de tumeurs basal-like [68], confirmant ainsi que 80% des tumeurs *BRCA1* sont de phénotype basal-like [91;96;97]. D'autres études du transcriptome des tumeurs *BRCA1* ont montré que celui-ci était enrichi en gènes impliqués dans le cycle cellulaire et la réponse immunitaire [98;99]. Un niveau supplémentaire d'hétérogénéité a été mis en évidence parmi les tumeurs *BRCA1* RE négatives, en fonction de l'expression faible ou forte du facteur de transcription NF κ B [99]. Les tumeurs *BRCA1* basal-like ont également été associées à un phénotype de stress oxydatif et d'hypoxie, mise en évidence par la surreprésentation du facteur HIF-1 α [100].

Une minorité de tumeurs *BRCA1* se caractérisent par l'expression du récepteur aux œstrogènes, une forte prolifération et un faible infiltrat lymphocytaire, lesquels sont des caractéristiques communes aux tumeurs de type luminale B. De manière intéressante, une étude datant de 2010, a montré que les femmes porteuses d'une mutation *BRCA1* développant un cancer exprimant les récepteurs aux œstrogènes, étaient plus âgées que celles développant un cancer de type basal-like [101].

1.3.3.1.2 Profil génomique des tumeurs BRCA1

Le génome des tumeurs *BRCA1* apparaît extrêmement remanié, ceci se traduisant par des altérations du nombre de copies et des mutations génétiques.

Les altérations du nombre de copies

L'analyse du génome des tumeurs *BRCA1* a largement bénéficié de l'avancée des technologies à la frontière entre cytogénétique et biologie moléculaire. Ainsi, Tirkkonen et al, ont été les premiers à utiliser la technique d'hybridation génomique comparative sur étalements métaphasiques, ou CGH chromosomique, pour analyser les anomalies de nombre de copies sur l'ensemble du génome tumoral. Dans leur étude, ils ont montré que les tumeurs *BRCA1* apparaissaient deux fois plus remaniées que les tumeurs sporadiques utilisées comme contrôles [102]. Les altérations les plus récurrentes étaient des pertes de régions chromosomiques dont le 5q, le 4q, le 4p, le 2q et le 12q.

Par cette même approche, l'équipe de Petra Nederlof a pu décrire une série d'anomalies (3p-, 3q- et 5q-) plus fréquemment observées dans des tumeurs *BRCA1* que dans des tumeurs de patientes à haut risque de cancer du sein. A l'aide de ces altérations génomiques, les auteurs ont construit un outil permettant d'identifier des patientes porteuses d'une mutation *BRCA1* avec une sensibilité de 96% et une spécificité de 76% [103;104]. Les auteurs ont ensuite confirmé leur outil de classification en augmentant la résolution de la technologie CGH sur puce BAC (pour *Bacterial Artificial Chromosome*) [105]. Les délétions des régions 5q et 4q ont également été retrouvées dans l'étude de Price et al, comparant le génome de tumeurs survenant dans un contexte familial par CGH chromosomique [106].

L'identification des sous-types tumoraux précédemment décrits, a contribué elle aussi au développement des signatures génomiques. En 2008, Melchor et al, ont montré que les différents profils génomiques correspondant aux différents sous-types tumoraux étaient retrouvés dans les cancers associés à une prédisposition héréditaire. Ils ont de plus démontré que le profil génomique des tumeurs *BRCA1* était proche de celui des tumeurs basal-like [107]. Plus récemment, deux études ont confirmé que les altérations précédemment identifiées comme spécifiques des tumeurs *BRCA1* étaient également retrouvées dans la signature génomique des tumeurs basal-like [108;109]. L'altération somatique du gène de prédisposition *BRCA1* dans les tumeurs sporadiques basal-like participe probablement à la grande similitude observée entre ces deux groupes de tumeurs [110].

Les amplifications de matériel génomique sont en revanche plus rares dans les tumeurs *BRCA1*. L'amplification de HER2 est effectivement peu fréquente et concerne moins de 15% des cas selon les études précédemment citées. La proximité du locus d'*ERBB2* à celui de *BRCA1*, perdu au cours de la tumorigénèse, a été proposée comme explication [111]. Le gène *MYC* est quant à lui amplifié dans plus de la moitié des cas selon Grushko et al [112] et l'amplification de la région 6q22-q24 contenant l'oncogène *MYB* a été décrite dans 30% des tumeurs *BRCA1* [113].

Par ailleurs, la mise en évidence de multiples copies de l'allèle muté de *BRCA1* a suggéré la possibilité d'une polyploïdie des tumeurs *BRCA1* [114]. Popova et al, se sont intéressés en 2012, à la ploïdie de ces tumeurs via le développement de la méthodologie GAP (pour *Genome Alteration Print*) reposant sur l'analyse de données de puces SNP [115]. L'innovation principale de cette méthodologie tenant à la représentation des profils de nombre

de copies et de statut allélique dans un graphique à deux dimensions, a permis de clarifier les profils génomiques complexes avec l'évaluation du nombre absolu de copies, l'identification de statuts particuliers comme les isodisomies, l'évaluation de l'hétérogénéité tumorale en termes de proportion de cellules tumorales et de sous-clones, et la détermination de la ploïdie sous-jacente. Grâce à cet outil, l'équipe de T. Popova et MH. Stern a pu démontrer que la majorité des tumeurs BRCA1 basal-like étaient pseudo-diploïdes (19/24) alors que les tumeurs sporadiques basal-like étaient pseudo-tétraploïdes (26/41; p <0,002), concluant ainsi que le statut pseudo-diploïde plaidait en faveur d'une inactivation du gène BRCA1 avec une valeur prédictive positive de 80% (IC 95% = 64% - 96%) [116]. Cette même équipe a également montré par l'étude sur puces SNP de 65 tumeurs de type basal-like, que les tumeurs BRCA1 (44/44) se distinguaient des tumeurs sporadiques (2/21) par un taux élevé de LST (p <10⁻¹²); les LST (pour Large-scale state transitions) se définissant comme des cassures chromosomiques entre des régions adjacentes d'au moins 10 Mb et correspondant le plus souvent à des translocations inter-chromosomiques [116]. Ce résultat a été validé ($p < 10^{-6}$) sur une deuxième série indépendante de 55 tumeurs basal-like et 17 lignées cellulaires basal-like. Parmi elles, 4 tumeurs et 2 lignées avec un taux élevé de LST présentaient une mutation de BRCA2. L'ensemble de ces données a permis aux auteurs de conclure que cette signature génomique était prédictive d'une inactivation de BRCA1/2 dans les tumeurs basal-like avec une sensibilité de 100% et une spécificité de 90%. Plus généralement, selon une étude publiée cette année, cette signature permettrait la détection de cancers du sein à recombinaison homologue déficiente par inactivation de l'un des facteurs majeurs impliqués dans cette voie, ouvrant la perspective de thérapies ciblées et de tests génétiques chez les patientes concernées [117].

Les altérations des gènes TP53 et PTEN

Les tumeurs *BRCA1* se caractérisent par une grande fréquence de mutations du gène *TP53*, d'abord mise en évidence par immunohistochimie alors qu'en condition normale, la protéine P53 n'est pas visible par immunomarquage du fait de sa courte demi-vie [89;118]. Des études de séquençage direct ont ensuite confirmé un excès de mutation de *TP53* dans les tumeurs *BRCA1* par rapport aux tumeurs sporadiques [119-121].

Le statut du gène *PTEN* a également été étudié et a fait l'objet d'une publication importante en 2008. L'équipe de Ake Borg a effectivement montré que l'expression fréquemment diminuée de PTEN dans les tumeurs de type basal-like s'accompagnait de mutations ponctuelles dans les tumeurs sporadiques, tandis qu'elle était associée à des réarrangements complexes dans les tumeurs *BRCA1*. Les auteurs ont ainsi suggéré que ces anomalies particulières étaient le signe d'un défaut de réparation de l'ADN du fait de l'absence de BRCA1 [122].

1.3.3.2 Caractéristiques des tumeurs du sein liées à une mutation constitutionnelle de *BRCA2*

Les cancers du sein liés à une mutation constitutionnelle de *BRCA2* sont majoritairement des tumeurs de type carcinome canalaire infiltrant de haut grade (grade II ou III). De nombreuses études rapportent également que la différenciation tubulaire des tumeurs *BRCA2* est faible (score de différenciation de 3), le pléomorphisme nucléaire modéré et le score mitotique élevé et significativement supérieur au groupe contrôle [80;123;124].

Les tumeurs *BRCA2* présentent majoritairement une positivité aux récepteurs hormonaux ainsi qu'une absence de surexpression des protéines ERRB2 et des cytokératines basales CK5/6 [89;124;125]. Ces critères classent les cancers du sein liés à *BRCA2* dans le phénotype « luminal » malgré leur haut grade histologique [108;124]. Elles sont également plus souvent mutées pour *TP53* que les tumeurs sporadiques et de nombreuses études ont mis en évidence la surexpression de protéines impliquées dans la régulation du cycle cellulaire et dans la carcinogénèse comme la Cycline D1 [126] et MYC [127].

1.3.3.2.1 Profil transcriptomique des tumeurs *BRCA2*

Les études transcriptomiques identifient les tumeurs *BRCA2* comme appartenant essentiellement au groupe luminal B de la classification intrinsèque d'expression de Sorlie [67].

En effet, Waddell et son équipe ont observé que 40% des tumeurs *BRCA2* étudiées en transcriptome (n=30) étaient luminales B et 33% étaient luminales A. Cette même étude a établi une signature prédictive des tumeurs *BRCA2* par la technique des centroïdes de shrunken, dont la validation a été insuffisante pour être utilisée en clinique [109]. L'étude d'Hedenfalk en 2001, a listé 11 gènes dont l'expression était significativement différente entre les tumeurs *BRCA2* et les tumeurs non *BRCA2* (tumeurs *BRCA1* et sporadiques). La

validation réalisée sur le même échantillonnage par un test de permutation randomisé a révélé seulement 4% de mauvaise classification. Cependant, cette signature n'a pas été validée sur un set de tumeurs indépendant [94]. Il est à noter que ces deux signatures des tumeurs *BRCA2* proposées par Hedenfalk et Waddell n'avaient aucun gène en commun.

Rouault et al, ont proposé en 2012 une signature basée sur les données d'expression permettant de distinguer les tumeurs *BRCA2* des tumeurs BRCAx, sur une cohorte de 7 tumeurs *BRCA2* et 24 tumeurs BRCAx ; les tumeurs BRCAx étant des tumeurs du sein survenues dans un contexte héréditaire pour lesquelles aucune mutation des gènes de prédisposition majeurs *BRCA1* et *BRCA2* n'a été caractérisée. Le *clustering* hiérarchique des 66 gènes de la signature, regroupant les échantillons en fonction de l'intensité d'expression des gènes, a permis de séparer les tumeurs à la fois sur le set d'entraînement et le set de validation, pour les classer en deux groupes : les tumeurs liées à *BRCA2* et les tumeurs BRCAx [128].

1.3.3.2.2 Profil génomique des tumeurs BRCA2

L'instabilité génétique qui caractérise les tumeurs *BRCA2* a conduit à rechercher des altérations génomiques, ou un profil d'altérations génomiques, propres à ces tumeurs.

Les études par CGH-Array ont permis d'établir l'existence de trois remaniements régulièrement associés: les délétions des bras longs des chromosomes 13 et 14 et les gains voire les amplifications géniques du bras long du chromosome 17 [102;108;129-131]. Ces résultats ont conduit le groupe de Petra Nederlof à proposer une signature prédictive des tumeurs *BRCA2* utilisant les centroïdes de shrunken qui, sur un groupe de validation de 38 tumeurs (19 mutées et 19 témoins), présentait une sensibilité de 89% et une spécificité de 84% [132].

La perte isolée des bras longs des chromosomes 13 et 14 a été décrite chez des contrôles dans certaines études précédemment citées. En revanche, l'association de perte des bras longs des chromosomes 13 et 14 semble renforcer le caractère spécifique de ce remaniement. Effectivement, une approche supervisée étudiant les profils génomiques d'une série de 9 tumeurs *BRCA2* et 9 tumeurs BRCAx, a retrouvé de manière spécifique aux tumeurs *BRCA2*,

une association de perte de matériel génétique au niveau des bras longs des chromosomes 13 et 14 [128]. Les régions retrouvées délétées dans les tumeurs *BRCA2* de cette étude, localisées entre les régions 13q13.3-13q14.3 et 14q24.2-14q32.2, ont d'ailleurs été confirmées par plusieurs études indépendantes [102;108;129-132]. Une étude plus récente encore, a également présenté la co-occurrence de la perte des régions 13q13 et 14q32 comme un marqueur génomique de la mutation *BRCA2* avec une spécificité de 90% et une sensibilité de 87% observées sur 29 tumeurs *BRCA2* et 130 tumeurs de haut-grade contrôles de statut *BRCA2* négatif ou inconnu [133].

L'analyse pathologique et moléculaire actuelle des cancers du sein, avec l'établissement d'une signature particulière, interroge sur l'existence d'une prédisposition héréditaire au cancer sous-jacente par mutation constitutionnelle des gènes *BRCA1* ou *BRCA2* et ainsi, sur l'accès potentiel au diagnostic génétique des patientes sans histoire familiale de cancer du sein.

1.4 Prédisposition héréditaire au cancer du sein

1.4.1 Les formes familiales de cancer du sein et leur déterminisme mendélien

Historiquement, l'idée de l'existence d'une composante héréditaire dans le cancer du sein remonte à 1757 où un chirurgien français, nommé Le Dran, raconte l'histoire d'une collègue ayant présenté un cancer du sein à l'âge de 19 ans et dont la grand-mère et un grand-oncle maternels étaient tous deux décédés de l'évolution de cette pathologie [134;135]. Un siècle plus tard, en 1866, Paul Broca décrit un regroupement anormal de cancers du sein au sein d'une famille, cette pathologie ayant touché sur quatre générations 10 femmes sur 24. Les données médicales et familiales récupérées par Paul Broca étaient alors suffisantes pour dresser un arbre généalogique et montrer la nature héréditaire de la maladie [136]. En 1971, Knudson propose un modèle de prédisposition génétique au cancer postulant que deux évènements mutationnels sont nécessaires au développement tumoral avec pour les formes héréditaires, une première mutation constitutionnelle héritée ou de novo sur un allèle, suivie d'une seconde mutation somatique sur l'autre allèle, dans une même cellule [137]. Ce modèle
découvert initialement dans le cas du rétinoblastome, s'applique depuis aux syndromes de prédisposition héréditaire au cancer en général, par la mutation constitutionnelle d'un gène de prédisposition au cancer.

Le rôle prépondérant des facteurs génétiques dans l'augmentation du risque de cancer du sein a été mis en évidence par l'étude de sœurs, et plus particulièrement par l'étude de sœurs jumelles [138;139]. En effet, le risque de développer un cancer du sein est augmenté pour des sœurs jumelles monozygotes dont l'une est atteinte, en comparaison avec des jumelles dizygotes ou les sœurs d'une fratrie.

Des études de ségrégation réalisées à partir des années 1980 ont cherché à définir le meilleur modèle génétique permettant d'expliquer les formes familiales de cancer du sein [52;140-142]. En 1991, l'une d'elles, portant sur 4730 cas de cancers du sein diagnostiqués chez des femmes de 20 à 54 ans et sur 4688 cas contrôles de même âge et localisation géographique, a permis de conclure à l'existence de deux populations parmi les personnes atteintes de cancer du sein : l'une sans évidence de déterminisme héréditaire correspondant à 95% des cas et l'autre minoritaire, constituée des 5% restants, en rapport avec la transmission d'un allèle de prédisposition de transmission autosomique dominante, de pénétrance élevée (avec un risque tumoral mammaire de l'ordre de 70% à 70 ans) et de faible fréquence (0,33%) [52].

Ce modèle mendélien, utilisé comme paramétrage des études de liaison, a été précieux pour permettre la localisation des gènes *BRCA1* [143] et *BRCA2* [144], respectivement en 17q21 et 13q12, puis leur identification par la mise en évidence de mutations responsables [145;146]. Ces découvertes ont constitué une avancée majeure dans la compréhension de la prédisposition héréditaire au cancer du sein, permettant de proposer des solutions de prévention aux femmes prédisposées génétiquement.

Il est cependant rapidement apparu qu'en dehors des situations de prédisposition de type seinovaire [147], l'implication des gènes *BRCA1* et *BRCA2* ne pouvait être retenue que pour une partie des cas de récurrence familiale de cancer du sein [148]. L'expérience des laboratoires de génétique moléculaire assurant les recherches de mutation des gènes *BRCA1* et *BRCA2* montre effectivement que lorsqu'est réalisée une recherche de mutation ponctuelle et une recherche de grand réarrangement, une altération des gènes *BRCA1* et *BRCA2* n'est identifiée que dans 20% (en 2004) ou 11% (en 2014) environ des familles explorées [149]. Les raisons de ce faible taux de mise en évidence de mutation pour des familles pourtant sélectionnées peuvent être multiples : (i) il peut s'agir d'altérations génétiques des gènes *BRCA1* et *BRCA2* non identifiables par les méthodes utilisées, (ii) l'analyse familiale peut être réalisée chez une personne correspondant en fait à une phénocopie, (iii) certains variants dont la signification pathologique reste inconnue peuvent en fait être délétères, et enfin (iv) certains cas familiaux correspondent probablement à des associations de cas sporadiques. La diminution du taux de détection des mutations entre 2004 et 2014 témoignent quant à elle de l'élargissement des critères de prescriptions des analyses génétiques adoptés en amont des laboratoires et interrogent sur la pertinence de certains d'entre eux.

Toutes ces raisons cependant ne permettent pas de rendre compte de la totalité des formes familiales de cancer du sein restantes. D'autres gènes que *BRCA1* et *BRCA2* sont donc impliqués dans d'assez nombreuses formes familiales de cancer du sein, appelées BRCAx [150]. Cette hypothèse a été explorée par liaison génétique dans une étude menée par le Breast Cancer Linkage Consortium à propos de 237 familles comportant au moins 4 cas de cancer du sein. Dans cette étude, si sur les 94 familles incluant des cas de cancer de l'ovaire, 81% étaient liés à *BRCA1* et si sur les 26 familles incluant des atteintes masculines, 76% étaient liés à *BRCA1* et *BRCA2*, dans 19 % des 34 familles d'au moins 6 cas de cancer du sein et dans 67% des 83 familles comportant 4 ou 5 cas de cancer du sein [151]. Ce chiffre de 67% corrobore le taux d'absence de mise en évidence de mutation observé en pratique clinique.

L'hypothèse de l'existence d'un autre mode de transmission pour les prédispositions BRCAx a été explorée. Deux études de ségrégation ont été conduites chez des patientes pour lesquelles aucune mutation de *BRCA1* et *BRCA2* n'avait été mise en évidence : l'une portant sur 824 familles australiennes, était en faveur d'un gène récessif qui concernerait une femme sur 250 [152], l'autre portant sur 1480 femmes anglaises atteintes de cancers du sein avant l'âge de 55 ans dont 21 porteuses d'une mutation *BRCA1* ou *BRCA2*, était également compatible avec l'existence d'un modèle récessif concernant 5 % des cas [153]. La possibilité d'un modèle polygénique a été également soulignée [153-156] et acceptée comme une explication pertinente des prédispositions au cancer du sein résiduelles [157].

Certains cas particuliers représentés par les syndromes de Li-Fraumeni, Peutz-Jeghers et Cowden que nous détaillerons plus tard (paragraphe 1.4.3.1.2), sont également associés à un excès de risque de cancer du sein. Ils sont cependant très rares et n'expliquent qu'une minorité de cas héréditaires de cancer du sein.

1.4.2 Les stratégies d'identification des facteurs génétiques de prédisposition au cancer du sein

Trois principales stratégies d'analyse ont été utilisées dans le passé pour identifier sur le plan moléculaire des facteurs génétiques prédisposant au cancer. Il s'agit des analyses de liaison génétique, des études de re-séquençage et des études d'association (Figure 2). Ces dernières années, suite au développement considérable des techniques de séquençage de l'ADN, c'est davantage par une approche de séquençage haut débit que de nouveaux gènes ou facteurs génétiques sont recherchés et identifiés [157;158]. L'ensemble de ces stratégies est présenté ci-après.

1.4.2.1 Les analyses de liaison génétique et le clonage positionnel

Les études de liaison génétique étudient la coségrégation de marqueurs génétiques avec un phénotype particulier dans le but de localiser des gènes en rapport avec un phénotype ou une maladie donné(e). Ces analyses, développées dans les années 1980, sont fondées sur l'étude de la ségrégation de gènes selon leur position sur un chromosome en observant la transmission d'un haplotype constitué d'un allèle de chacun des loci d'intérêt. Si deux loci sont proches sur le génome, l'haplotype se transmet comme un seul gène, en revanche si les loci sont éloignés ou situés sur deux chromosomes différents, on observera une ségrégation indépendante des loci. Les études de liaison nécessitent pour réussir, d'analyser de nombreux membres d'une ou plusieurs grandes familles afin de distinguer clairement les individus atteints et les individus non atteints pour le phénotype étudié. De même, elles sont pertinentes uniquement pour la localisation de gènes associés à une pénétrance élevée. Dans ce cas, les mutations génétiques sont fortement associées au phénotype et la majorité des individus atteints en seront porteurs. S'il s'agit d'un gène associé à une faible pénétrance, la corrélation entre le phénotype et le statut mutationnel peut être insuffisant pour générer un signal de liaison. La mesure statistique utilisée pour déterminer la probabilité que deux loci soient liés

est appelée LOD score (pour *logarithm of odds score*) ; un LOD score de +3 ou plus, étant en faveur d'un lien fort ne pouvant être attribué au hasard et un LOD score de -2 ou moins, étant en faveur d'une absence de liaison [159;160]. Une fois qu'une région génomique candidate est signalée par cette approche, le clonage positionnel peut ensuite être utilisé pour affiner la région candidate jusqu'à ce que le gène d'intérêt ou les mutations en cause soient identifiés [161].

Ces études de liaison, par l'analyse de grandes familles de cancer du sein, ont conduit à la localisation en 1990 et 1994, des deux gènes majeurs de prédisposition héréditaire au cancer du sein que sont les gènes *BRCA1* au locus 17q21 [143] et *BRCA2* au locus 13q12-13 [144], ainsi qu'à leur clonage quelques années plus tard, respectivement en 1994 et 1995 [145;146]. Des études de liaison ont également été réalisées dans des groupes de familles présentant un phénotype particulier pour cartographier les loci spécifiques en lien avec des syndromes comportant un risque élevé de cancer du sein. Le clonage positionnel a ensuite été utilisé pour identifier les gènes en cause, tels que *PTEN*, *STK11*, et *CDH1* (Figure 2) [162-167].

Cependant, le cancer du sein est de nature complexe et les études de liaison ne sont pas toujours simples à réaliser. Un élément clé des études de liaison est effectivement d'obtenir des échantillons de tous les membres d'une famille, et en particulier de ceux qui sont atteints. Ces derniers pouvant être décédés ou inaccessibles pour d'autres raisons, il est parfois impossible de recueillir leur prélèvement sanguin. De plus, le cancer du sein est une maladie hétérogène avec différents phénotypes et pouvant être affectée par des facteurs environnementaux. L'ensemble de ces facteurs a pu limiter la découverte d'autres gènes de prédisposition au cancer du sein par cette méthode.

En effet, quelques loci candidats ont été rapportés ces dernières années : la région 8p12 suggérée par l'analyse de 2 familles distinctes, ou la région 13q21 sur des résultats de CGH chromosomique à propos de 61 tumeurs appartenant à 37 familles BRCAx [168] mais dans les deux cas, ces loci ont été exclus par étude de liaison génétique sur 128 familles BRCAx [169;170]. Le locus 17q11-q12 a également été proposé sur des approches de CGH chromosomique, de LOH (*Loss of Heterozygosity*) et de liaison génétique à propos de 18 tumeurs familiales BRCAx, mais aucune mutation n'a pu être caractérisée sur dix gènes candidats de la région [171].



Figure 2: Relation entre la pénétrance de la maladie, la fréquence allélique et les stratégies d'identification des facteurs génétiques de prédisposition au cancer du sein (d'après Aloraifi et al, Surgical Oncology 2015 [158]).

1.4.2.2 Les études dites de re-séquençage ou le criblage mutationnel de gènes candidats

Les études de re-séquençage impliquent le criblage mutationnel de gènes sur des échantillons d'individus partageant un phénotype commun. L'acquisition d'une meilleure compréhension de l'étiologie moléculaire du cancer du sein a permis dans les années 1990, de proposer un aperçu des gènes candidats possiblement impliqués dans la prédisposition au cancer du sein.

Il semblait biologiquement intéressant et plausible, que les protéines interagissant avec BRCA1 et BRCA2 ou agissant dans les voies similaires de réparation de l'ADN puissent être impliquées dans la prédisposition au cancer du sein [172]. Les mutations délétères touchant les gènes impliqués dans la réparation de l'ADN sont cependant très rares car majoritairement éliminées par le phénomène de sélection naturelle et nécessitent l'étude d'un grand nombre de cas et de témoins pour permettre une comparaison statistiquement significative entre les deux groupes. Cette approche exige de plus le criblage de l'ensemble de la séquence codante du gène afin de détecter de nouvelles mutations responsables.

Des expériences de cette ampleur ont été publiées pour une poignée de gènes impliqués dans la réparation de l'ADN et souvent associés à une pénétrance intermédiaire tels que *CHEK2*, *ATM* et *PALB2* (Figure 2) [173-175]. D'autres gènes candidats pressentis comme ceux impliqués dans la régulation du cycle cellulaire, les points de contrôle, l'apoptose et le métabolisme des hormones stéroïdiennes, ont rarement été évalués dans cette perspective [157].

Ainsi, les études de liaison génétique ont échoués à identifier de nouveaux gènes de susceptibilité et l'identification de variants rares de gènes impliqués dans la réparation de l'ADN n'explique qu'un très petit nombre de cas.

1.4.2.3 Les études d'association

Au cours de ces vingt dernières années, l'idée que les maladies complexes résultaient en fait d'un effet combiné de plusieurs allèles communs conférant un risque faible a émergé.

Malgré le succès des premières études de liaison pour identifier des variants responsables dans les maladies mendéliennes, celles-ci manquaient effectivement de puissance pour détecter des mutations pathogènes impliquées dans les maladies complexes, causées par de multiples variants génétiques à effet réduit en combinaison avec des facteurs environnementaux. Pour répondre à cette contrainte et grâce à l'émergence de nouvelles techniques de génotypage à grande échelle par puces SNP (pour *Single Nucleotide Polymorphism*), de larges études d'association ont été développées et étendues au génome entier : les GWAS (pour *Genome-Wide Association Studies*) (Figure 2) [176].

Les études d'association analysent la fréquence d'un variant donné dans une population d'étude versus la population témoin dans le but d'identifier des associations entre un génotype particulier et un phénotype, en l'occurrence ici la survenue d'un cancer du sein. Ces études reposent en fait sur l'hypothèse qu'il existe des segments chromosomiques partagés entre les populations humaines parce qu'ils ont été hérités d'un ancêtre commun. Ces segments partagés ont été brisés par recombinaison à travers de nombreuses générations et sont désignés par des blocs haplotypiques [177]. En comparant la fréquence de milliers de SNP chez des milliers d'individus atteints et d'individus témoins, des associations peuvent être déduites et se traduire par la localisation dans le génome d'un variant à risque. Cette association dépend du déséquilibre de liaison ; le déséquilibre de liaison se définissant comme l'association non aléatoire des allèles à deux ou plusieurs loci, suggérant qu'ils puissent être physiquement proche et donc liés. Les études d'association vont donc générer une masse importante de données, due à la quantité de marqueurs utilisés et à l'importance des effectifs.

La grande nouveauté de ces études par rapport aux études précédentes, réside dans le fait que les populations étudiées ne sont plus forcément des familles avec une récurrence de cancers du sein, mais des patientes provenant de la population générale. Elles sont de plus dites « sans a priori » vis-à-vis d'un mode de transmission. Enfin, ce type d'étude est parfaitement adapté à l'identification de variants fréquents prédisposant au cancer du sein dans la mesure où il est plus facile de mettre en évidence des différences statistiquement significatives entre un groupe d'individus atteints et un groupe de témoins, lorsque la fréquence des variants est supérieure à 5% dans la population générale. Les résultats de ces études vont ainsi consister en l'observation de variants génétiques, codants ou non, communément présents dans la population générale et reliés à une faible augmentation du risque de cancer du sein.

C'est dans ce but que des centaines de milliers de SNP communs ont été génotypés pour tester leur co-occurrence avec le cancer du sein. Les premières études ont porté sur des gènes choisis pour leur fonction protéique, examinant ainsi l'association qu'il pouvait y avoir entre de multiples variants concernant des gènes candidats reconnus dans la prédisposition et le cancer du sein. Cependant, ces expériences étaient souvent de petite taille, sujettes à des biais et utilisaient des niveaux de significativité trop indulgents, entraînant une incohérence et un manque de réplication des résultats. Ainsi, la majorité des associations supposées correspondaient plus vraisemblablement à des faux positifs [178;179]. Afin d'augmenter la puissance des études et rationaliser le choix des variants candidats, des collaborations ont été réalisées pour entreprendre des études d'association sur des dizaines de milliers d'échantillons. Depuis, de nombreuses GWAS ont été menées sur le cancer du sein et ont permis de valider de nombreux variants qui seront discutés plus tard dans le paragraphe 1.4.3.3 de l'introduction : « Les allèles de prédisposition fréquents associés à une pénétrance faible ».

1.4.2.4 Les études de séquençage massif haut débit

Ces dernières années, les améliorations spectaculaires des techniques de séquençage de l'ADN, devenues plus rapides, plus efficaces et moins couteuses, ont offert une alternative réaliste aux approches par GWAS pour identifier de nouveaux gènes de prédisposition et en particulier des variants rares, à risque élevé de cancer du sein [180].

Le terme de « séquençage de nouvelle génération » ou NGS (pour *Next Generation Sequencing*) est utilisé pour décrire le séquençage en parallèle de millions de courtes séquences. Il est différent du séquençage Sanger traditionnel qui utilise la séparation par électrophorèse capillaire des produits d'ADN précédemment amplifiés par PCR (pour *Polymérase Chain Reaction*) [181;182]. Au lieu de séquencer de longs fragments d'ADN, la technique NGS consiste à fragmenter l'ADN pour obtenir de courtes séquences qui pourront ensuite être alignés sur le génome de référence. Trois grandes méthodes NGS ont été développées en fonction du coût des études et de l'efficacité recherchée, à savoir : le séquençage ciblé, le séquençage d'exome et le séquençage du génome entier [183;184].

Le séquençage de type exome se concentre sur la capture d'environ 30 mégabases d'exons [185] tandis que le séquençage du génome entier a pour but de séquencer plus de 3 milliards de paires de bases, comprenant environ 50% de séquences ADN répétées [186]. Cette dernière

technologie, en se soumettant à de plus grandes régions d'intérêt, est moins précise en raison d'une profondeur de lecture (nombre de séquences générées) plus faible. Les séquences répétées compliquent de plus l'alignement des séquences au génome de référence du fait d'une grande similitude entre elles.

Depuis 2009, les analyses d'exomes ont permis une meilleure connaissance des bases moléculaires des maladies génétiques, même à partir d'un nombre restreint de patients [187;188]. Ces études ont utilisé des stratégies d'analyse comparative d'exomes. En 2013, la conférence annuelle de l'American Society of Human Genetics (ASHG) a montré que le WES (pour Whole Exome Sequencing) était déjà utilisé pour les applications diagnostiques de maladies génétiques dans certains pays. L'expérience d'une équipe américaine a rapporté, chez 250 enfants présentant pour la plupart un phénotype neurologique anormal sans diagnostic, l'identification d'une altération génétique chez 25 % des patients grâce au séquençage de l'exome [189]. Ces auteurs étaient capables, grâce au WES, d'identifier l'altération du gène en cause dans des maladies génétiques aux modes de transmission différents que ce soit autosomique dominant, autosomique récessif ou lié à l'X. De la même façon, ils ont détecté différents types de mutations : faux sens, non sens, insertions et délétions avec ou sans décalage du cadre de lecture et mutations d'épissage. Ce type d'approche a été appliquée dans le domaine de la prédisposition héréditaire au cancer du sein et a permis d'identifier de potentiels nouveaux gènes de prédisposition tels que XRCC2, FANCC, BLM et plus récemment RECQL, mais ces derniers semblent n'intervenir chacun que dans un petit nombre de familles [190-192].

Le séquençage ciblé permet de capturer sélectivement des régions génomiques d'intérêt, soit en solution, soit par des sondes moléculaires d'inversion, qui peuvent ensuite être amplifiées. Grâce à cette technologie, une énorme quantité de données peut être générée par région, offrant ainsi une grande confiance dans l'identification des variants. Effectivement, par rapport au séquençage de l'exome, le séquençage de gènes cibles nécessite moins de régions séquencées par individu permettant d'augmenter la profondeur de lecture de chaque base. En outre, par un système de code-barres (ou barcoding) des ADNs fragmentés, plusieurs échantillons peuvent être regroupés afin de séquencer plusieurs échantillons en une seule expérience. La démocratisation de cette technologie permettant de réduire considérablement son coût ainsi que sa rapidité et sa capacité à identifier tout type de mutations a favorisé son introduction dans le diagnostic moléculaire des maladies génétiques dont les syndromes de prédisposition héréditaire au cancer. Le développement de diverses méthodes pour capturer des régions génomiques spécifiques d'intérêt a permis de se concentrer sur des cohortes particulières de gènes ou des régions de gènes pour un séquençage complet.

Deux équipes ont ainsi évalué, en 2010 puis 2014, les performances de cette technologie pour l'analyse des gènes *BRCA1* et *BRCA2* et d'autres gènes ayant un rôle probable dans les cancers du sein et/ou de l'ovaire chez des patients porteurs d'une mutation connue [193;194]. Les auteurs ont développé deux sets de capture ciblant les régions codantes, non codantes, introniques ainsi que 10 kb en amont et en aval des gènes *BRCA1* et *BRCA2* associés à 19 gènes pour Walsh et al et à 26 gènes pour Castera et al. Les deux études ont montré que cette nouvelle technologie était capable de détecter des mutations ponctuelles, des insertions et délétions et des réarrangements géniques avec une sensibilité estimée à 100 %. A noter que l'analyse des gènes autres que *BRCA1* et *BRCA2*, est aujourd'hui réalisée dans les laboratoires au titre de la recherche mais devrait permettre une meilleure connaissance de leur déterminisme dans les tumeurs du sein et/ou de l'ovaire. De plus, l'enrichissement en séquences non codantes offre un réservoir d'informations qui pourra être utilisé ultérieurement, comme une étape intermédiaire avant l'étude du séquençage portant sur le génome entier.

Actuellement, des études de séquençage testant des panels de gènes impliqués dans le cancer du sein sont utilisés, tels que le panel BreastNext (Ambry genetics) ou le panel BROCA (Université de Washington) précédemment évoqué et dont le nom fait référence à Paul Broca et à son Traité des tumeurs [193]. Ceux-ci ont d'ailleurs été appliqués récemment sur une série de 1150 individus, permettant d'identifier *PPM1D* comme potentiel gène candidat prédisposant au cancer du sein avec la détection de variants en mosaïque [195].

Ainsi, le séquençage de nouvelle génération, quelques soit l'approche utilisée, a permis de grandes avancées dans le diagnostic et la compréhension des maladies génétiques. Son application au diagnostic en routine nécessite cependant d'être encadrée pour éviter des dérives d'interprétation même après une analyse comparative d'exomes. Le véritable challenge de l'analyse d'exome voire du génome, reste l'interprétation des variants détectés, les données générées étant considérables et de manipulation difficile [196].

1.4.3 Les gènes de prédisposition héréditaire au cancer du sein

Les facteurs génétiques identifiés à ce jour et impliqués dans l'augmentation du risque de cancer du sein sont classés en fonction de leur pénétrance (probabilité de présenter la maladie quand on est porteur de la mutation) et de leur fréquence dans la population générale. On distingue ainsi trois grandes classes de facteurs génétiques : (i) les mutations rares à forte pénétrance (risque élevé de cancer du sein), (ii) les mutations rares à pénétrance modérée (risque intermédiaire de cancer du sein) et (iii) les variants fréquents à pénétrance faible (risque faible de cancer du sein) (Figure 2) [157]. A noter que la notion de risque élevé se réfère à des variants avec un risque relatif (RR) supérieur à 5, alors que la notion de risque intermédiaire et de risque faible se réfère respectivement à des variants avec un RR compris entre 2 et 5 et un RR inférieur à 2.

1.4.3.1 Les gènes de prédisposition associés à une pénétrance élevée

Deux gènes majeurs, *BRCA1* et *BRCA2*, sont associés à un risque très élevé de cancer du sein (RR > 10) lorsqu'ils sont altérés constitutionnellement. L'implication d'autres gènes à pénétrance élevée, a été proposée dans la prédisposition héréditaire au cancer du sein. Il s'agit des gènes *TP53*, *PTEN*, *CDH1* et *STK11*, chacun de ces gènes étant responsable d'un syndrome particulier associé à un risque de cancer du sein bien que celui-ci ne soit pas au premier plan du syndrome. Ces différents gènes sont présentés ci-après.

1.4.3.1.1 Les gènes majeurs : BRCA1 et BRCA2

Les gènes *BRCA1* et *BRCA2*, sont considérés comme les gènes majeurs de la prédisposition héréditaire au cancer du sein. Ils sont localisés respectivement en 17q21 et 13q12-13 et ont été découverts entre les années 1990 et 1995 grâce à une approche classique de localisation génique par étude de liaison génétique et clonage positionnel, précédemment décrite dans le paragraphe 1.4.2.1 [143-146]. Ce sont des gènes de grande taille ayant des structures génomiques proches, notamment avec des exons géants. Ils ne présentent cependant aucune homologie de séquence. Le gène *BRCA1* s'étend sur 81 kilobases (kb) et il est constitué de 24 exons, dont 22 sont codants. Le gène *BRCA2* s'étend sur 84 kb et il est constitué de 27 exons, dont 26 sont codants. Tous deux possèdent un exon 11 de très grande taille qui représente 61% de la séquence codante pour *BRCA1* et 48% pour *BRCA2*. Ils codent deux grandes

protéines, constituées de 1863 pour BRCA1 et 3418 acides aminés pour BRCA2, qui ne présentent ni homologie de séquence, ni homologie de structure entre elles ou avec d'autres protéines.

Structure et fonctions des protéines BRCA1 et BRCA2

La protéine BRCA1 est une protéine ubiquitaire, dont la localisation est principalement nucléaire. Elle possède plusieurs domaines fonctionnels dont un domaine en doigt de zinc de type RING (pour *Really Interesting New Gene*) en N-terminal, un domaine central de liaison à l'ADN et deux domaines BRCT (pour *BRCA1 C-Terminus*) en tandem, situés en C-terminal. Elle contient également deux signaux de localisation nucléaire, codés par l'exon 11 et situés entre les résidus 341 et 748 [197], un domaine coiled-coil situé entre les résidus 1393 et 1424 permettant la fixation de la protéine PALB2 [198] ainsi que deux signaux d'export nucléaire (Figure 3).

La répétition en tandem des domaines BRCT semble impliquée dans les interactions protéineprotéine de BRCA1 avec d'autres protéines comme BRCA2, RHA (*RNA Helicase A*), CtIP (*CtBp-interacting protein*) et HDAC (*Histone déacetylase*) 1 et 2 [199;200]. Le domaine RING est impliqué dans l'ubiquitination, processus multi-étape faisant intervenir successivement trois grandes classes d'enzymes appelées E1-, E2- et E3-ligases, permettant l'adressage d'un substrat modifié vers le protéasome pour sa dégradation [201]. Il est le siège de l'hétérodimérisation entre BRCA1 et BARD1 (*BRCA1 Associated RING Domain 1*), qui possède elle aussi un domaine RING et deux domaines BRCT. En l'absence de BARD1, BRCA1 est instable et rapidement dégradé [202]. Le recrutement de la protéine BARD1 permet notamment l'interaction du domaine RING finger de BRCA1 avec l'ADN. Le domaine RING fixe également la protéine BAP1 (*BRCA1-associated protein 1*), une enzyme présentant une activité d'ubiquitine-hydrolase et impliquée dans la prédisposition au cancer [200]. Les domaines BRCT et RING sont particulièrement conservés dans l'évolution et l'observation fréquente de mutations constitutionnelles dans ces domaines, associées à une prédisposition au cancer du sein, démontre leur importance dans la fonction de BRCA1 [203].

Plusieurs fonctions de BRCA1 ont été élucidées et ont permis de comprendre son implication dans la tumorigénèse (pour revues [204;205]). En effet, ces rôles incluent la réparation de l'ADN (par recombinaison homologue ou excision de nucléotide), le contrôle du cycle



Figure 3 : Structures et protéines partenaires de BRCA1 et BRCA2 (adapté de Welcsh et al, Trends Genet 2000 [200]).

cellulaire, la régulation transcriptionnelle de gènes cibles tels que *MDM2*, *BAX*, *P21/WAF et GADD45*, le remodelage de la chromatine via des interactions avec les protéines SWI/SNF et BACH1, ainsi que le maintien du chromosome X inactif par le recrutement du transcrit *XIST* (*X inactive Specific Transcript*) [206]. L'association de BRCA1 à BARD1 permet la participation de ce complexe à tous ces processus grâce à ses fonctions de facteur de transcription, ainsi qu'à son action d'ubiquitine ligase E3.

La protéine BRCA2 possède en ce qui la concerne, un domaine de transactivation en Nterminal, 8 domaines BRC répétés (pour *BRCA2-repeats*) situés dans la partie centrale et codés par l'exon 11 ainsi qu'un signal de localisation nucléaire en C-terminal (Figure 3) (pour revues [204;205]). Elle est principalement impliquée dans la réparation des cassures doublebrin de l'ADN par recombinaison homologue [200]. En effet, BRCA2 interagit via ses domaines BRC avec la protéine clé de la recombinaison homologue, RAD51, et régule la fixation de celle-ci au niveau du site de cassure [207;208]. A noter par ailleurs que la protéine PALB2 contrôle la localisation de BRCA2 et maintient sa stabilité au niveau de la chromatine. Elle semble ainsi être critique pour les fonctions de BRCA2 corrélées à la réparation de l'ADN et à l'activation du point de contrôle en phase S [209].

Les protéines BRCA1 et BRCA2 jouent donc un rôle important dans le maintien de la stabilité du génome en permettant notamment une réparation précise et efficace des cassures doublebrin de l'ADN. BRCA2 semble essentiellement impliquée dans la réparation par recombinaison homologue alors que BRCA1 intervient dans différents processus cellulaires, présentant ainsi un rôle plus large et en amont de BRCA2. De manière intéressante, de nouvelles molécules appelées inhibiteurs de PARP ou anti-PARP parce qu'elles inhibent l'action de PARP1 (pour poly(ADP-ribose) polymérase 1), exploitent ces défauts de réparation par recombinaison homologue induits par la perte fonctionnelle de BRCA1 ou BRCA2 [210;211]. L'enzyme PARP1 est une ADN polymérase qui contribue au maintien de l'intégrité de l'ADN en éliminant les cassures simple brin ou les bases endommagées par un système de réparation, le système BER (pour Base excision repair). L'absence de ce mode de réparation dans des cellules déficientes en BRCA1 ou BRCA2 ne peut dans ce cas être compensée par le recours à la réparation des cassures double-brin induites lors de la phase S à l'emplacement des cassures simple brin, entraînant de manière sélective la mort cellulaire selon un mécanisme de synergie létale ou « synthetic lethality ». Ces molécules, associées à une moindre toxicité, ont depuis montré leur efficacité dans ce contexte mutationnel [212].

Actuellement, l'une d'elles (l'Olaparib) a obtenu l'AMM dans le traitement des cancers de l'ovaire en rechute, sensibles aux sels de Platines et d'autres, voire la même molécule, sont testées dans le traitement des cancers du sein métastatiques ou triple négatifs (essais cliniques de phase III de type BRAVO ou OLYMPIA).

Prédisposition héréditaire au cancer liée aux gènes BRCA1 et BRCA2

Les gènes *BRCA1* et *BRCA2* sont des gènes suppresseurs de tumeur répondant au modèle de Knudson avec l'inactivation biallélique du gène de prédisposition par mutation constitutionnelle d'un allèle puis inactivation somatique du second allèle dans les cellules cancéreuses [137]. Les mutations les concernant sont donc des mutations de type perte de fonction ou inactivatrices.

De nombreuses mutations constitutionnelles ont été décrites à leur niveau (plus de 6000 variants différents répertoriés pour les deux gènes dans la base BRCA ShareTM) dont des mutations fondatrices, fréquentes dans certaines populations comme les mutations c.185delAG (*BRCA1*), c.5382insC (*BRCA2*) et c.6174delT (*BRCA2*) chez les juifs Ashkénazes et la mutation c.999del5 (*BRCA2*) chez les icelandais.

Dans la majorité des cas cependant, il s'agit de mutations individuellement rares et beaucoup ont été rapportées dans seulement quelques familles voire une seule. La plupart d'entre-elles sont des mutations dites tronquantes car interrompant le cadre de lecture. Il s'agit de mutations non sens, de délétions et insertions de nucléotides avec décalage du cadre de lecture et de mutations d'épissage.

Un grand nombre de mutations faux sens et synonymes sont également rapportées pour les deux gènes. La plupart sont considérées comme bénignes ou neutres et d'autres, comme des variants de signification inconnue pour lesquels les données actuelles ne permettent pas de statuer sur leur pathogénicité. Quelques mutations faux sens cependant, par exemple celles concernant les résidus cystéine du domaine RING de BRCA1, abolissent la fonction de la protéine et confèrent un risque similaire aux mutations tronquantes [213;214].

Il existerait de plus, une corrélation génotype-phénotype pour les mutations *BRCA1* et *BRCA2*, avec l'observation d'un nombre de cancers de l'ovaire par rapport au nombre de

cancers du sein plus élevé lorsque les mutations surviennent dans la région centrale des deux gènes plutôt qu'aux extrémités 3' et 5' [215;216]. Les mécanismes biologiques sous-tendant ces observations restent cependant assez nébuleux [217].

Les mutations des gènes *BRCA1* et *BRCA2* sont associées à une majoration importante du risque de cancer qui est essentiellement mammaire et ovarien chez la femme. En effet, le risque cumulé à 70 ans de développer un cancer du sein est de 65% pour *BRCA1* et 45% pour *BRCA2* alors qu'il est d'environ 10% dans la population générale [218]. Le risque relatif de cancer du sein lié à ces deux gènes semble de plus évoluer différemment avec l'âge. En effet, le RR de cancer du sein est majeur avant l'âge de 40 ans pour les porteuses de mutation *BRCA1* puis ce RR diminue avec l'âge. Pour les porteuses de mutation *BRCA2*, ce RR augmente progressivement avec l'âge, comme dans la population générale mais à un niveau plus élevé. Concernant le cancer de l'ovaire, le risque cumulé à 70 ans de le développer est de 39% pour *BRCA1* et 11% pour *BRCA2* alors qu'il est d'environ 1% dans la population générale [218]. Ce risque est lui aussi plus précoce en cas de mutation *BRCA1* (< 50 ans) qu'en cas de mutation *BRCA2* (> 50 ans). Les mutations *BRCA1* et *BRCA2* rendent ainsi compte de la plupart des formes familiales de cancer du sein et de l'ovaire, notamment lorsque de nombreux cas de chaque sont rapportés dans les familles [219-222].

Plus modestement, les mutations *BRCA* et en particulier les mutations *BRCA2*, exposent à d'autres risques tumoraux, comme le cancer du pancréas (RR = 3,51) et de la prostate (RR = 4,65), notamment avant l'âge de 65 ans pour ce dernier (RR = 7,33) [223;224]. Le risque de cancer du sein masculin est également plus élevé en cas de mutation *BRCA2* avec un risque cumulé à 70 ans estimé à 6% [225].

Enfin, des mutations bialléliques du gène *BRCA2* ont été identifiées et sont responsables d'un sous-type particulier de l'anémie de Fanconi, le sous-type D1 [226]. L'anémie de Fanconi est un syndrome rare de transmission autosomique récessive se caractérisant par une instabilité chromosomique, un syndrome malformatif congénital, l'apparition progressive d'une aplasie médullaire ainsi qu'une prédisposition aux hémopathies et aux tumeurs solides. Le sous-type D1 se distingue des autres sous-types par une aplasie médullaire non systématique et un risque élevé et précoce de leucémie et de tumeurs solides spécifiques comme le médulloblastome et le nephroblastome [227]. Les mutations bialléiques du gène *BRCA1*, sont en revanche plus exceptionnelles, la combinaison des deux mutations étant en général létale à

l'état embryonnaire [228]. A ce jour, elles n'ont été rapportées que chez deux patientes et dans les deux cas, l'une des deux mutations était une mutation faux sens [229;230].

1.4.3.1.2 Les autres gènes à pénétrance élevée: TP53, PTEN, STK11 et CDH1

Les mutations des gènes *TP53* (gène de prédisposition au syndrome de Li-Fraumeni) et *STK11* (gène de prédisposition au syndrome de Peutz-Jeghers) confèrent un risque très élevé de cancer du sein, avec un risque cumulé compris entre 40 et 90% [49]. Les mutations des gènes *PTEN* (gène de prédisposition au syndrome de Cowden) et *CDH1* (gène de prédisposition au cancer gastrique de type diffus) sont elles aussi associées à risque élevé de cancer du sein avec un risque cumulé compris entre 25 et 60% [49]. Les mutations de ces quatre gènes sont responsables de syndromes de prédisposition au cancer rares, en général facilement reconnaissables car ils présentent des caractères particuliers.

Le gène TP53

Le syndrome de Li Fraumeni (LFS) est une affection rare, de transmission autosomique dominante, prédisposant à diverses tumeurs dont le cancer du sein pré-ménopausique, le sarcome des tissus mous et l'ostéosarcome, le corticosurrénalome, les tumeurs cérébrales, les hémopathies malignes, et bien d'autres tumeurs encore [231]. Ce syndrome se caractérise en outre par des atteintes précoces, voire pédiatriques [232] et une hypersensibité aux radiations ionisantes avec un risque de tumeurs secondaires radio-induites important [233]. Il résulte de l'altération constitutionnelle du gène suppresseur de tumeur *TP53*.

La rareté de l'affection et le tableau clinique dramatique associé avec une mortalité importante notamment chez les enfants, ont exclu la possibilité d'étude de liaison pour ce syndrome. Il est cependant apparu très tôt que *TP53* était le siège d'altérations génétiques fréquentes en pathologie tumorale humaine [157;234], qu'il était un facteur de transcription central impliqué dans de multiples voies de signalisation cellulaire et jouait un rôle crucial dans le contrôle du cycle cellulaire [235]. Il permet en outre l'arrêt du cycle cellulaire en cas de dommage à l'ADN, l'activation des protéines de réparation de l'ADN et induit l'apoptose lorsque la réparation est impossible.

C'est donc en 1990, par une approche « gène candidat », que des mutations causales de *TP53* ont été identifiées dans cinq familles [236], celles-ci exposant à des risques élevés de cancer, notamment mammaire. En effet, le risque de développer un cancer en cas de mutation *TP53* est environ de 15% à 15 ans, de 80% pour les femmes à 50 ans et de 40% pour les hommes au même âge [237]; la différence significative observée entre les sexes étant presque totalement expliquée par le cancer du sein [237;238].

Les cancers du sein survenant dans un contexte de mutation *TP53* ont pour particularité d'être souvent très précoces (< 30 ans) et de surexprimer l'Oncogène Her2 [239-241]. La probabilité d'identifier une mutation *TP53* chez une femme ayant développé un cancer du sein avant l'âge de 30 ans n'est cependant que de 4 à 8% en l'absence d'histoire personnelle et/ou familiale évocatrice d'un syndrome de Li-Fraumeni [238;242-244].

Le gène STK11

Le syndrome de Peutz-Jeghers est une polypose hamartomateuse de l'ensemble du tube digestif, qui s'accompagne d'une lentiginose labile de la muqueuse buccale, de la sphère anale et des doigts. Il s'agit d'un syndrome rare dont l'incidence est estimée entre 1/50000 à 200000 individus [245].

Les pertes d'hétérozygotie observées dans les polypes d'individus atteints du syndrome de Peutz-Jeghers ont aidé à établir en 1997 une liaison pour 12 familles sur le chromosome 19p, conduisant rapidement en 1998 à l'identification du gène en cause par clonage positionnel [163;164]. Le gène *STK11*, également appelé *LKB1* (pour *Liver kinase B1*) est un gène suppresseur de tumeur. Il code en effet pour une sérine/thréonine kinase contrôlant la polarité cellulaire, inhibant la prolifération cellulaire et interagissant avec la voie de signalisation TOR [246].

Les mutations constitutionnelles du gène *STK11* confèrent un risque élevé de cancer pour le tube digestif et la glande mammaire mais également pour le pancréas, les ovaires (cellules de la granulosa), les testicules (cellules de Sertoli), le col utérin et les poumons, nécessitant la mise en place d'une surveillance spécifique [245]. Le risque de développer un cancer du sein en cas de syndrome de Peutz-Jeghers est effectivement élevé avec un risque cumulé compris entre 30 et 50% [49;247;248].

Le gène PTEN

Le gène *PTEN* (pour *Phosphatase and tensin homology deleted on chromosome ten*), code une phosphatase à double spécificité (sérine et thréonine) jouant un rôle de suppresseur de tumeur en régulant négativement la voie de signalisation cellulaire PI3K/AKT/mTOR [249;250]. Ses mutations sont responsables de la maladie de Cowden, un syndrome autosomique dominant très rare dont l'incidence est estimée à 1/200000 individus [251].

C'est en 1996, par une étude de liaison génétique entreprise à propos de 12 familles présentant une maladie de Cowden, qu'une liaison génétique a pu être établie entre les marqueurs de la région chromosomique 10q22-10q23 et le locus de la maladie [252]. A la même époque, il était noté qu'une altération fréquemment retrouvée dans de nombreux types tumoraux était la perte d'hétérozygotie au locus 10q23.3. Cette donnée ainsi que la suppression de la tumorigénicité de cellules de glioblastome chez la souris par la réintroduction d'un allèle du locus 10q23.3, suggéraient que ce locus contenait un gène suppresseur de tumeur [253]. C'est ainsi qu'en 1997, trois équipes différentes identifiait en même temps le gène *PTEN* [254-256]. Les données de clonage positionnel du gène *PTEN* individualisant un locus très exactement superposable à celui qui avait été défini pour la maladie de Cowden, il devenait un candidat majeur pour être responsable de cette maladie. L'analyse de *PTEN* chez les patients présentant une maladie de Cowden a effectivement très vite mis en évidence la présence de mutations de ce gène à l'état hétérozygote [257].

Cliniquement, la maladie de Cowden est de présentation complexe. Elle se manifeste notamment par une macrocéphalie quasi constante et la survenue au cours de l'existence de lésions hyperplasiques, hamartomateuses ou tumorales touchant de multiples organes dont la peau et les muqueuses, la glande mammaire, la thyroïde et le tube digestif. Son diagnostic est parfois difficile du fait de la grande variabilité d'expression qui la caractérise. On regroupe désormais sous le terme de PHTS (pour *PTEN Hamartoma Tumor Syndrome*) [258], les différents syndromes cliniques caractérisés par une altération constitutionnelle du gène *PTEN*, tels que la maladie de Cowden, le syndrome de Bannayan-Riley-Ruvalcaba (le versant pédiatrique de la maladie), la maladie de Lhermitte-Duclos, le syndrome de macrocéphalie et autisme, certaines formes de polypose juvénile et d'autres manifestations d'hypercroissance segmentaire. La complication majeure de cette affection est représentée par la survenue de tumeurs malignes touchant principalement le sein et la thyroïde avec respectivement, un

risque cumulé à 70 ans de 77% (IC 95% = 59% - 91%) et 38% (IC 95% = 25% - 56%) [259]. Le risque mammaire est donc très élevé en cas de mutation *PTEN*, et finalement assez proche de celui occasionné par des mutations *BRCA1* et *BRCA2*. D'autres risques tumoraux ont été rapportés dans la maladie de Cowden mais dans une moindre mesure pour l'endomètre, le rein et la peau [259].

Le gène CDH1

Une étude de liaison génétique menée en Nouvelle Zélande, sur une très grande famille Maori où ségrégeaient de multiples cas de cancers gastriques de type diffus, a permis d'identifier le gène *CDH1* comme gène responsable [260;261]. Ce gène code l'E-Cadhérine, une protéine transmembranaire importante dans le maintien de la polarité cellulaire et l'adhésion des cellules entre-elles dans le tissu épithélial [262].

Depuis, de nombreuses familles de cancers gastriques liées à *CDH1* ont été rapportées et une fréquence élevée de cancer du sein de type lobulaire a pu être constatée [263;264]. Les données de pénétrance associées aux mutations *CDH1* ont été mises à jour récemment à partir de l'étude de 75 familles. D'après cette étude, le risque cumulé à 80 ans de développer un cancer gastrique diffus chez les porteurs de mutation *CDH1* est estimé à 70% pour les hommes et à 56% pour les femmes. Celles-ci présentent en plus un risque élevé de cancer du sein lobulaire avec un risque cumulé à 80 ans de 42% [265]. Plus occasionnellement, des mutations *CDH1* ont été décrites dans des formes familiales de cancers du sein lobulaires sans cas de cancer gastrique associé [266].

L'implication des gènes *TP53*, *STK11*, *PTEN* et *CDH1* dans la prédisposition héréditaire au cancer du sein est évidente. Ils sont d'ailleurs les seuls aujourd'hui, avec les gènes *BRCA1* et *BRCA2*, à faire l'objet d'un diagnostic moléculaire chez les patientes susceptibles de présenter une prédisposition génétique car l'identification d'une mutation délétère à leur niveau permet de proposer une prise en charge spécifique et pluridisciplinaire, adaptée aux risques tumoraux. Les syndromes dont ils sont responsables sont cependant rares et ne rendent compte que d'une faible proportion des cancers du sein héréditaires (<5%) [150]. De plus, les autres éléments du syndrome sont en général au premier plan et les mutations de ces gènes, si elles prédisposent au cancer du sein, ne sont pas ou peu responsables de simples récurrences familiales de ce

type de tumeurs (références pour *TP53* [267-269], *STK11* [270], *PTEN* [271-273] et *CDH1* [266]).

1.4.3.2 Les gènes de prédisposition associés à une pénétrance modérée

La mise en évidence du fait que les altérations constitutionnelles bialléliques de *BRCA2* étaient responsables d'un des groupes de complémentation de l'anémie de Fanconi a conduit à explorer les différents gènes impliqués dans ce syndrome et plus généralement, les gènes codant pour les protéines partenaires de BRCA1 et BRCA2 et impliquées dans la réparation de l'ADN [172]. Plusieurs gènes de prédisposition à pénétrance intermédiaire ont ainsi été identifiés via une approche « gène candidat ». Il s'agit des gènes *ATM*, *CHEK2*, *BRIP1* et *PALB2* ainsi que les gènes sont associées à un risque relatif de cancer du sein compris entre 2 et 4. Elles sont rares avec une fréquence estimée dans la population générale entre 0,1 et 0,5% et retrouvées dans environ 4% des cancers du sein familiaux.

Le gène ATM

Le gène *ATM* (pour *Ataxia-telangiectasia mutated*) a été localisé en 1988 sur le chromosome 11q22-23 par analyse de liaison, puis identifié en 1995 par clonage positionnel chez des patients présentant un syndrome d'Ataxie-Télangiectasie [274;275]. Il s'étend sur 150kb et est composé de 66 exons dont 62 codent pour une importante protéine kinase appartenant à la famille des PIKKs (pour *PI3 kinase-like protéine kinases*) et s'exprimant dans tous les tissus bien que plus fortement au niveau de la rate, du thymus, des testicules, des cellules de Purkinje cérébelleuses et des cellules endothéliales (cérébrales, cutanées et conjonctivales).

La protéine ATM possède comme l'ensemble des PIKKs, un domaine catalytique présentant des homologies avec le domaine catalytique de la PI3K (pour *Phosphatidyl Inositol-3 Kinase*) (Figure 4). Ce domaine est encadré par un domaine FAT en N-terminal (de 500 acides aminés) et un domaine FATC en C-terminal (de 35 acides aminés). La fonction exacte du domaine FAT est peu connue mais une des hypothèses retenue par le groupe d'Abraham est



Figure 4: Motifs communs aux PI3KKs (d'après Y. Shiloh, Nature Reviews Cancer 2003 [277]).



Figure 5: Rôle de la protéine ATM dans la réparation des cassures double-brin de l'ADN (d'après Choi et al, Molecular Cancer Therapeutics 2016 [284]).

que les domaines FAT et FATC interagiraient avec la structure tridimensionnelle de l'extrémité C-terminale des protéines et permettraient ainsi la bonne conformation du domaine catalytique, permettant la liaison de l'ATP et la phosphorylation du substrat [276]. De plus, c'est dans le domaine FAT que se trouve la serine 1981, site d'auto-phosphorylation de la protéine ATM permettant son activation [277]. La partie N-terminale est quant à elle constituée de nombreuses répétitions de motifs HEAT (*Huntingtin, Elongation factor 3, A subunit of protein phosphatase 2A, and TORI*). Ces répétitions permettent les interactions protéine-protéine et donc aux PI3KKs d'interagir avec leurs partenaires ainsi que leurs cibles. De plus, les domaines d'interaction protéine-protéine étant très éloignés du domaine catalytique en structure primaire, ce domaine N-terminal permet à la protéine d'adopter une conformation en super-hélice et ainsi de les juxtaposer en structure tertiaire [276].

La protéine ATM est impliquée dans de nombreuses fonctions cellulaires et aide au maintien de l'intégrité du génome en jouant un rôle central dans le contrôle du cycle cellulaire et dans la détection et la signalisation des dommages causés à l'ADN [277;278] (Figure 5). Elle intervient ainsi dans la réparation des cassures double-brin (CDB) de l'ADN, ces dernières pouvant être induites par des radiations ionisantes, des molécules de chimiothérapie, un stress oxydatif ou lors d'événements physiologiques normaux tels que la recombinaison méiotique ou la recombinaison V(D)J impliquée dans la synthèse des immunoglobulines. Deux mécanismes de réparation sont alors activés: la recombinaison homologue (RH) entre deux molécules sœurs d'ADN et la jonction non homologue des brins brisés (ou NHEJ pour non homologous end joining) [279]. La réparation des CDB implique un vaste réseau de protéines, jouant le rôle de capteur et médiateur du signal, de transducteur et d'effecteur [280;281]. Brièvement, lorsque ce type d'évènement survient, le complexe MRN (Mre11/Rad50/Nbs1) est recruté au site de l'ADN endommagé et recrute à son tour la protéine ATM. Celle-ci, présente normalement dans la cellule sous forme de dimère inactif, s'auto-phosphoryle au niveau de la sérine 1981, entraînant sa dédimérisation et son activation [282;283]. Elle phosphoryle ensuite H2AX, permettant le recrutement au site endommagé de plusieurs protéines effectrices activées par ATM, dont P53, BRAC1, CHK1 et CHK2. Ces dernières vont ensuite induire l'arrêt du cycle cellulaire et l'activation des points de contrôle pour permettre la réparation de l'ADN endommagé ou l'apoptose en cas de réparation impossible [282;284] (Figure 5). Il a depuis été démontré qu'une multitude de cibles (plus de 1000) étaient phosphorylées par ATM, notamment à la suite de CDB de l'ADN, et qu'ATM régulait d'autres réseaux de protéines non impliqués directement dans la réparation de l'ADN mais dans certaines voies métaboliques telles que la voie du récepteur à l'IGF1 (pour *insulin-like growth Factor 1*) ou de réponse au stress oxydatif [280;285-288]. Il a ainsi été rapporté qu'ATM jouait un rôle dans la synthèse d'antioxydants, notamment dans les cellules de Purkinje du cervelet, permettant de lutter contre le stress cellulaire et la neurodégénérescence [289;290].

A l'état homozygote ou le plus souvent hétérozygote composite, les mutations germinales du gène *ATM* sont responsables, du syndrome d'ataxie-télangiectasie (AT) (OMIM : <u>208900</u>), un syndrome autosomique récessif rare, dont la fréquence en population caucasienne est estimée entre 1/40000 à 1/300000 individus [278]. Cette maladie se caractérise entre autres, par un syndrome neurodégénératif avec une ataxie cérébelleuse débutant vers l'âge de deux ans, des télangiectasies oculo-cutanées, un déficit immunitaire avec des infections respiratoires graves, une instabilité chromosomique se traduisant par des translocations entre les chromosomes 7 et 14 visibles au caryotype lymphocytaire, et une prédisposition au cancer chez l'enfant, essentiellement à type d'hémopathies malignes [291-293]. Plus de 300 mutations distinctes ont été décrites et s'observent tout le long du gène, sans site prédominant de mutations [278;294]. La plupart sont des mutations tronquantes entraînant une perte d'expression de la protéine ATM. Certaines mutations faux-sens ont cependant été rapportées mais ne représentent que 10% environ des mutations *ATM* identifiées chez les patients AT [278].

A l'état hétérozygote, les mutations *ATM* concerneraient entre 0,5 et 1% de la population occidentale [174;295]. Or, elles semblent prédisposer elles aussi à la survenue de certains cancers dont et principalement le cancer du sein [296;297]. En effet, dès 1976, Swift et al rapportaient un excès de cancers du sein chez les apparentées féminines de patients atteints d'AT [298]. Cette observation alors que le gène *ATM* n'était pas encore identifié, s'est vérifiée des années après dans plusieurs grandes études épidémiologiques de familles AT, retrouvant un risque augmenté mais modéré de cancer du sein chez les hétérozygotes, avec un RR compris entre 2 et 4 selon les études [296;297;299-302]. Du fait de ces données, plusieurs investigateurs ont conduit des études de criblage mutationnel sur des séries de cancers du sein sporadiques dans le but de clarifier le rôle des mutations *ATM* dans la prédisposition au cancer du sein. Les résultats obtenus étaient pour la plupart contradictoires et ont finalement échoué à démontrer son implication [299;303-309]. Des arguments tels que la taille des effectifs, le choix des populations étudiées et les techniques de criblage utilisées à l'époque, ont été proposés comme explication à cette absence de résultats [310;311]. Cependant, une plus large

étude menée en 2006 et portant cette fois sur 443 cas de cancer du sein familiaux et 521 individus contrôles, a réussi à démontrer que le risque de cancer du sein était augmenté chez les hétérozygotes *ATM* non AT, mais ce risque restait modéré, avec un RR à 2,37 (p = 0,0003) [174], proche de celui observé dans les grandes études épidémiologiques de familles AT [302].

En pathologie sporadique, il est apparu que le gène *ATM* était fréquemment altéré, particulièrement dans les hémopathies malignes mais également dans de nombreux autres types tumoraux. Près de 180 mutations tronquantes et 500 mutations faux sens différentes sont rapportées dans la base COSMIC (pour *Catalogue Of Somatic Mutations In Cancer*) et sont observées tout le long du gène [234]. L'expression d'ATM est par ailleurs fréquemment réduite dans les cancers du sein sporadiques et une perte d'hétérozygotie au locus d'ATM est observée dans environ 40% des cas [312]. De manière intéressante, il a été montré récemment qu'à l'instar des tumeurs *BRCA1* et *BRCA2*, une expression faible d'ATM dans les cellules cancéreuses mammaires, leur conférait une sensibilité aux inhibiteurs de PARP [313], laissant suggérer que les indications de ce type de traitement pouvaient s'étendre aux patientes porteuses d'une mutation *ATM*.

Le gène CHEK2

Le gène *CHEK2*, localisé sur le chromosome 22q12.1, code une autre protéine kinase jouant un rôle central dans la médiation de la réponse aux dommages causés à l'ADN. Elle est ellemême activée par ATM en cas de cassure double-brin de l'ADN pour activer à son tour d'autres protéines clés impliquées dans la réparation et le contrôle du cycle cellulaire telles que P53 et BRCA1 [314].

Les mutations de *CHEK2*, et particulièrement la mutation c.1100delC, ont été pour la première fois impliquées dans la prédisposition héréditaire au cancer en 1999, par Bell et al, dans des familles plutôt évocatrices d'un syndrome de Li-Fraumeni, laissant suggérer que *CHEK2* pouvait être un nouveau gène de prédisposition à ce syndrome [315]. Cependant, les mutations de *CHEK2* ont montré qu'elles étaient trop fréquentes dans la population générale (environ 1%) pour être en cause dans le syndrome de Li-Fraumeni.

En revanche, l'identification quelques années après par Meijers-Heijboer et al, de la mutation c.1100delC dans 4,2% (30/718 ; p = 5 x 10⁻⁶) de formes familiales de cancer du sein, a montré qu'il pouvait être un gène de susceptibilité au cancer du sein de pénétrance intermédiaire avec un RR estimé à 1,70 (IC 95% = 1,32 – 2,20) [316]. En combinant les données de dix études cas-témoins, le consortium *CHEK2* a ensuite démontré que la mutation c.1100delC était retrouvée dans 1,9% (201/10860) des cas de cancer du sein comparé à 0,7% (64/9065) chez les témoins (p = 1x10⁻⁷), avec un odds ratio à 2,34 (IC 95% = 1,72 – 3,20) [317].

Il a également été proposé que cette mutation pouvait être associée à d'autres risques tumoraux, notamment prostatiques, mais les données sont à ce sujet contradictoires [316;318-321]. Par ailleurs, d'autres mutations tronquantes de *CHEK2* ont été rapportées dans des cas de cancer du sein ainsi que des mutations faux sens, telle que la mutation c.470T>C – p.Ile157Thr, mais la signification pathogénique de ces dernières reste à ce jour mal établie [322-327].

Le gène BRIP1

Le gène *BRIP1* (aussi appelé *BACH1*), code une hélicase qui interagit avec la protéine BRCA1 et joue de ce fait un rôle dans la réparation de l'ADN et le contrôle du cycle cellulaire [328;329]. C'est en étudiant les protéines interagissant avec BRCA1 que Cantor et al, ont rapporté des mutations germinales de *BRIP1* dans deux cas précoces de cancer du sein, suggérant une implication de ce dernier dans la prédisposition au cancer du sein [328].

En 2006, des mutations tronquantes de *BRIP1* ont effectivement été retrouvées chez 9/1212 femmes *BRCA1/2* négatives, atteintes de cancer du sein à caractère familial, contre 2/2081 chez les témoins ; ces résultats confortant l'hypothèse que *BRIP1* était un gène de prédisposition au cancer du sein (p = 0,003), de pénétrance intermédiaire car associé à un RR de 2 (IC 95% = 1,2 – 3,2 et p = 0,012) [330]. Une autre étude réalisée sur une cohorte australienne et néo-zélandaise retrouvait également une mutation tronquante de *BRIP1* parmi 75 cas de cancers du sein héréditaires et aucune chez les 93 témoins testés [331]. Cependant, d'autres études menées sur des cohortes différentes (chinoise, américaine, franco-canadienne, finlandaise, italienne, suisse et allemande), n'ont détecté aucune mutation tronquante de *BRIP1* chez les patients présentant une apparente prédisposition familiale [332-340]. De manière intéressante, il est apparu que des mutations bialléliques de *BRIP1* étaient responsables d'un des groupes de complémentation de l'anémie de Fanconi, le sous-groupe J. Ce dernier est cependant de présentation phénotypique différente du sous-groupe D1 lié aux mutations bialléliques de *BRCA2*, notamment car il n'a pas été associé à des tumeurs solides chez l'enfant [341-343].

Le gène PALB2

Le gène *PALB2* (pour *Partner and Localizer of BRCA2*) code une protéine permettant la localisation et la stabilisation de BRCA2, l'aidant ainsi dans ses fonctions, notamment de réparation des cassures double-brin de l'ADN par recombinaison homologue [344].

En 2006, une équipe travaillant sur *PALB2* montrait que son inactivation rendait les cellules sensibles aux cassures chromosomiques par action de la mitomycine C [344], phénomène classiquement observé dans l'anémie de Fanconi et conduisant une année plus tard à la détection de mutations bialléliques de *PALB2* dans cette affection, responsables elles-mêmes du sous-groupe N [345]. Le phénotype associé à ce sous-groupe s'apparentait d'ailleurs à celui causé par les mutations bialléliques de *BRCA2* avec une fréquence plus élevée de tumeurs solides chez l'enfant de type médulloblastome et néphroblastome (ou tumeur de Wilms) que dans les autres sous-groupes [346].

A l'instar des mutations *BRCA2* et *BRIP1* impliquées dans l'anémie de Fanconi à l'état biallélique et dans la prédisposition héréditaire au cancer du sein à l'état monoallélique, *PALB2* est apparu comme un gène candidat de prédisposition au cancer du sein. En 2007, Rahman et al, ont mené une large étude cas-témoins portant sur des formes familiales de cancers du sein et ont effectivement identifié une mutation tronquante de *PALB2* dans 10/923 cas alors qu'aucune mutation n'a été trouvée chez les 1084 témoins testés (p = 0,0004). Le risque relatif de cancer du sein pour les hétérozygotes *PALB2* a pu être estimé à partir de l'analyse de ces familles à 2,3 (IC 95% = 1,4 - 3,9; p = 0,0025) [175].

Il a été montré plus récemment que les mutations de *PALB2* conféraient un risque de cancer du sein proche de celui des mutations *BRCA2*, avec un risque cumulé à 70 ans de 33% en l'absence d'antécédent familial de cancer du sein (IC 95% = 25 - 44) et de 58% (IC 95% = 50

- 66) lorsque l'histoire familiale de la patiente était démonstrative avec au moins deux cas de cancer du sein rapportés chez ses apparentés avant l'âge de 50 ans [347].

Des mutations fondatrices ont par ailleurs été rapportées en Finlande (c.1592delT) et au Canada (c.2323C>T) [348;349].

Les gènes du complexe MRN

Le complexe MRN est un complexe protéique hautement conservé, constitué par trois protéines : RAD50, MRE11A et NBS1. Lorsqu'une cassure double-brin survient dans l'ADN, ce complexe est recruté au site endommagé et recrute à son tour la protéine ATM comme précédemment décrit. Ce complexe intervient donc dans la réparation des cassures doublebrin de l'ADN mais également dans la recombinaison méiotique, le contrôle du cycle cellulaire et le maintien des télomères [350;351].

Une étude finlandaise a montré qu'une mutation tronquante du gène *RAD50* (c.657delT) était retrouvée 4 fois plus fréquemment dans les cas de cancer du sein (8/317) que chez les témoins (6/1000) (p = 0,008; OR = 4,3; IC 95% = 1,5 – 12,5), suggérant qu'il s'agissait d'une mutation fondatrice non retrouvée dans les autres cohortes de pays nordiques [352]. D'autres mutations tronquantes de *RAD50* ont été rapportées mais la contribution de ce gène dans la prédisposition héréditaire au cancer du sein en dehors de la Finlande nécessite d'autres investigations [353].

Le gène *NBS1* (ou *NBN*) constitue lui aussi un gène de susceptibilité modéré avec un risque de cancer du sein augmenté, 2 à 3 fois supérieur à celui de la population générale, chez les hétérozygotes (notamment porteurs de la mutation c.657del5) [354]. Les mutations homozygotes du gène *NBS1* sont quant à elles responsables du syndrome de Nijmegen, se caractérisant par un syndrome dysmorphique et polymalformatif avec macrocéphalie et retard de croissance, un retard intellectuel de survenue progressive, un déficit immunitaire avec des infections des voies respiratoires sévères à répétition, ainsi qu'une forte prédisposition aux tumeurs malignes (surtout lymphoïdes) et une radiosensibilité (OMIM : <u>251260</u>).

De la même manière, le gène *MRE11A*, localisé sur le chromosome 11q21, est apparu comme un potentiel gène candidat de prédisposition héréditaire au cancer du sein avec l'observation

en 2008, de deux mutations le concernant sur une série de huit familles avec récurrence de cancer du sein non liées à *BRCA1/2* mais associées à une perte d'expression des protéines du complexe MRN en tissus tumoral [355]. A l'état homozygotes ou hétérozygote composite, les mutations du gène *MRE11A* sont responsables d'un syndrome proche de l'AT, appelé ATLD (pour *Ataxia-telangiectasia-like disorder*) [356]. Ce dernier se présente comme une forme moins sévère d'AT, cependant sans télangiectasie, avec un début plus tardif et une progression plus lente de la maladie que dans la forme classique. En revanche, les patients ATLD ne semblent pas présenter de prédisposition au cancer d'après le peu de cas rapportés [292].

Plus généralement, une étude cas-témoins américaine menée en 2014 sur des cas de cancer du sein de survenue précoce (< 45 ans), a démontré que des mutations dans les gènes du complexe MRN étaient retrouvées dans 2,2% (29/1312) des cas de cancer du sein comparé à 0.6% (7/1121) chez les témoins, attestant que *RAD50, NBS1 et MRE11A* constituaient des gènes de prédisposition au cancer du sein de pénétrance intermédiaire, avec un odds ratio à 2.88 (p = 0.0090). La contribution de ces gènes dans la prédisposition héréditaire au cancer du sein reste cependant faible au vu du nombre de mutations identifiées, la majorité étant de plus des mutations faux sens [357].

Le gène *RAD51C*, a également été rapporté ces dernières années, comme gène candidat de prédisposition héréditaire au cancer du sein et de l'ovaire dans une étude allemande ainsi que dans une étude finlandaise et une étude australienne, retrouvant des mutations de ce gène dans des cas de cancer de l'ovaire ou dans des familles présentant un syndrome sein-ovaire, mais aucune dans les familles présentant uniquement une agrégation de cas de cancer du sein [358-360]. Cependant, deux études récentes, l'une espagnole et l'autre pakistanaise, ont détecté chacune une mutation inactivatrice de *RAD51C* dans une famille présentant uniquement une récurrence de cas de cancers du sein [361;362]. Le risque de cancer du sein associé aux mutations de *RAD51C* n'est donc pas encore clairement établi.

1.4.3.3 Les allèles de prédisposition fréquents associés à une pénétrance faible

Il est apparu ces vingt dernières années que la prédisposition aux maladies communes résultait de l'effet combiné de multiples allèles associés chacun à un risque faible (RR < 1,5) [363]. Ce type de variant contribue très probablement au cancer du sein selon un modèle polygénique

modifiant la prédisposition héréditaire au cancer du sein. Comme précédemment évoqué dans le chapitre 1.4.2.3, ces variants ont été découverts par l'intermédiaire des études d'association ou GWAS. Ces dernières ont révélés l'existence de nombreux SNPs à fréquence allélique élevée significativement associés à la survenue de cancers du sein.

Les premiers SNPs identifiés en 2007 par trois équipes différentes et montrant la plus forte association (avec des OR compris entre 1,07 et 1,28), étaient situés dans ou à proximité des gènes suivants: *FGFR2, TNRC9, MAP3K1, LSP1* et *H19* [364-366]. Un autre polymorphisme (rs1045485) identifié la même année par l'équipe de Cox et al, était lui aussi intéressant car il semblait à l'inverse des précédents, avoir un effet protecteur contre le cancer du sein (OR = 0,88 ; IC 95% = 0,84 – 0,92) et concernait le gène *CASP8*, lui-même codant pour la caspase-8, une protéine impliquée dans l'apoptose [367]. A la suite de ces divers résultats, des études ciblant spécifiquement certains de ces gènes ont été réalisées, afin d'en approfondir la connaissance. Ainsi, une étude des SNPs associés aux gènes *FGFR2* (rs2981582) et *TNRC9* (rs3803662) a rapporté une augmentation du risque de cancer du sein chez les porteurs d'une mutation de *BRCA2* et de l'allèle mineur du gène *FGFR2*, de même que chez les porteurs d'une mutation des gènes *BRCA1* ou *BRCA2* et de l'allèle mineur du gène *TNRC9* [368].

A ce jour, plus de 1000 études d'association sur le sujet ont été publiées ces vingt dernières années, impliquant plus de 7000 variants [369]. Les résultats de ces études ont montré une forte association pour les variants précédemment cités et ceux situés dans ou à proximité des gènes suivants: *NOTCH2, NEK10, MRSP30, ECHDC1, CDKN2A/B, ANKRD16, ZNF265, ZMIZ1, RAD51L1, TOX3, COS11, ANKLE1, CHEK2, ATM, NBN, CTLA4, ERCC2, ESR1, GSTM1, GSTT1, HSD17B, IFNG, IGF1, LRTOMT, MTHFR, NUMA1, TP53, TYMS, VDR, and WRN* [364-367;369-371]. Des études de réplication à grande échelle ont ainsi identifié 79 loci et une étude plus récente encore, datant de 2015, a permis d'identifier 15 loci supplémentaires dont l'association avec le cancer du sein était significative [370]. A noter que la plupart de ces études ont été menées sur des échantillons d'individus européens. Néanmoins, quelques SNPs montrant une association significative ont été trouvés parmi les populations asiatiques, dans ou à proximité des gènes suivants: *ESR1, GSTP1* et *XRCC3* [369].

De manière intéressante, un certain nombre de SNPs identifiés par les GWAS et significativement associés à un risque de cancer du sein, sont intergéniques tels que les loci

2q35, 8q24 et 11q13. Ils sont donc situés à distance de toute région codante et cette observation a pu susciter quelques doutes quant à la pertinence biologique des variants identifiés par les GWAS [372]. Il est à noter cependant que la plupart des variants intergéniques révélés par les GWAS (76,5% selon Maurano et al.) sont situés sur des sites d'hypersensibilité à la DNAse [373]. Or ces sites correspondent à des régions du génome où la chromatine est ouverte, exposant l'ADN et le rendant accessible. Ils sont donc particulièrement présents dans les régions de transcription et de régulation des gènes actifs et il a été suggéré que les variants situés dans ces sites, pouvaient agir comme des éléments cis-régulateurs de gènes particuliers et être impliqués dans l'étiologie d'une pathologie. Pour exemple, il a été démontré que le locus 8p24, également impliqué dans le cancer colorectal, altérait le site de fixation d'un facteur de transcription activant à distance l'oncogène *MYC* [374;375].

Finalement, ces études d'association ont permis l'identification de nombreux allèles liés à une prédisposition au cancer du sein. La pénétrance de ces allèles de susceptibilité reste cependant extrêmement faible et ils correspondent sans doute davantage à des gènes modificateurs dans le cadre d'une hérédité multigénique qu'à d'authentiques allèles de susceptibilité [376]. De manière intéressante, Milne et Antoniou ont étudié le risque cumulé de cancer du sein en fonction de la distribution de 18 SNPs connus pour être associés au risque de cancer du sein dans la population générale [377]. Cette analyse a été réalisée pour des femmes de la population générale ainsi que pour des femmes porteuses d'une mutation BRCA2. Dans le premier groupe, 5% des femmes considérées à faible risque (car non porteuses des SNPs à risque étudiés) avaient un risque cumulé attendu $\leq 5,7\%$ alors que 5% des femmes considérées à haut risque (car porteuses des SNPs à risque étudiés) avaient un risque cumulé attendu \geq 19%. Dans le groupe des mutées *BRCA2*, 5% des femmes considérées à faible risque avaient un risque cumulé attendu $\leq 47\%$ alors que 5% des femmes considérées à haut risque avaient un risque cumulé attendu \geq 89%. Les auteurs ont ainsi montré une majoration significative du risque de cancer du sein lié à ces 18 SNPs dans les deux groupes, avec un résultat plus marqué en cas de mutation BRCA2. En revanche, le même résultat n'a pas été observé pour les femmes porteuses d'une mutation BRCA1. Le principal argument avancé pour expliquer cette constatation est que la plupart des SNPs identifiés par les GWAS, l'ont été à partir de cas de cancer du sein sporadiques et sont donc plus fréquemment associés à des cancers du sein exprimant les récepteurs aux œstrogènes. Or, à la différence des tumeurs du sein BRCA2, les tumeurs du sein BRCA1 ne les expriment généralement pas.

2 OBJECTIFS DE TRAVAIL

La découverte des gènes *BRCA1* et *BRCA2* et la description de leurs altérations constitutionnelles ont conduit à la prise en compte sur un plan médical des situations de prédisposition héréditaire aux cancers du sein et de l'ovaire. Cependant, les recherches de mutation de ces deux gènes dans les situations de récurrence familiale de cancer du sein restent relativement décevantes dans la mesure où une altération délétère n'est détectée que dans 30 à 40 % des cas. Ainsi, plus de la moitié des suspicions cliniques de prédisposition héréditaire au cancer du sein restent aujourd'hui inexpliquées sur le plan mutationnel et les tentatives d'individualiser un autre gène majeur de prédisposition au cancer du sein (*BRCA3*) ont jusqu'à présent toujours échoué.

Une des raisons semble être l'importante hétérogénéité génétique qui s'observe lorsqu'un simple critère de récurrence familiale est utilisé pour sélectionner les familles étudiées. Effectivement, d'assez nombreux cas ponctuels de formes familiales de cancer du sein ont été rapportés à des mutations de gènes variés comme *BARD1*, *PALB2* ou *BRIP1* mais également *TP53* ou *PTEN*, l'implication de chacun de ces gènes étant très limitée en termes de nombre de familles concernées. L'identification des gènes en cause repose de ce fait sur l'étude de familles individuelles, rarement suffisamment étendues pour permettre une stratégie de localisation génique selon les méthodes classiques de liaison génétique ou d'association.

La stratégie envisagée ici consiste à combiner une approche constitutionnelle de séquençage d'exomes applicable à de très petits échantillons, à une approche tumorale par la recherche d'haplotypes communs dans les aberrations chromosomiques propres aux tumeurs ayant touché différents membres d'une même famille. En effet, le développement d'un cancer de forme familiale implique généralement (i) l'héritage d'un (ou plusieurs) facteur(s) de prédisposition (mutation, polymorphisme) provenant d'un ancêtre commun, et (ii) l'activation de ce facteur par une aberration chromosomique (typiquement la disparition de l'allèle sauvage par LOH dans le cas d'un gène suppresseur de tumeur mais également sa duplication ou son amplification dans le cas d'un oncogène). Le génotypage tumoral par puce SNP haute résolution a pour but d'identifier au sein d'une même famille, les segments haplotypiques communs qui subissent un même type d'aberration chromosomique. L'idée originale de ce

projet est que le jeu combiné des crossing-over et des remaniements génomiques tumoraux nous permettrait d'identifier des régions candidates de taille modérée au sein desquelles un gène de prédisposition pourrait être recherché par séquençage de nouvelle génération (ou NGS) et la mutation causale identifiée.

Une sélection de plusieurs familles avec récurrence de cancers du sein et disponibilité de l'ADN tumoral pour au moins deux membres de ces familles a ainsi fait l'objet de cette double approche.

3 MATÉRIEL ET MÉTHODES

3.1 Matériel

L'activité d'oncogénétique développée à l'Institut Bergonié depuis les années 1990 a conduit à un recrutement important de formes familiales de cancer du sein. L'informatisation dès le début de cette activité des généalogies concernées a permis d'avoir accès à un grand nombre de familles évocatrices d'une situation de prédisposition héréditaire au cancer du sein non liées à *BRCA1* et *BRCA2*. Le croisement de ces données avec le fichier de la tumorothèque de l'Institut Bergonié a ensuite permis de sélectionner les familles pour lesquelles au moins deux échantillons tumoraux cryo-préservés de cancer du sein étaient disponibles. 17 familles avec récurrence de cancers du sein ont ainsi été sélectionnées dont 14 avec un échantillon tumoral pour deux de ses membres et trois avec un échantillon tumoral pour trois de ses membres familiaux. A noter que l'ensemble de ces familles avait bénéficié au préalable d'un criblage mutationnel des gènes *BRCA1* et *BRCA2* avec identification d'une mutation délétère dans trois familles (familles G1995049 et G2003221 (mutation *BRCA1*) et famille G2009373 (mutation BRCA2)) servant de témoins positifs à l'étude. Les arbres généalogiques de ces familles sont présentés en annexe (Annexes 1 à 17).

Au total, nous disposions de 37 ADNs tumoraux provenant de 17 familles, dont 14 sans mutation *BRCA1/2* identifiée, pour réaliser les génotypages tumoraux sur puces SNP. Nous disposions également pour ces derniers de 27 ADNs constitutionnels (leucocytaires) appariés ainsi que 37 autres ADNs constitutionnels d'individus apparentés (Tableau 2).

Sept échantillons d'ADN leucocytaire ont ensuite été sélectionnés parmi les cas index des familles précédentes afin de réaliser un séquençage haut débit de type exome.

Enfin, 17 échantillons supplémentaires de cancers du sein, développés par 16 patientes différentes provenant de dix nouvelles familles, ont été sélectionnés pour réaliser un séquençage haut débit ciblé (sur un panel de gènes impliqués dans la prédisposition au cancer). Ces familles, dont les arbres généalogiques sont fournis dans l'article (Sup Figures 1

à 10), se caractérisaient toutes par une agrégation familiale de cancers du sein associée à une mutation constitutionnelle du gène *ATM* identifiée chez le cas index. A noter que tous les cas index de ces familles ont bénéficié au préalable d'une procédure de diagnostic génétique moléculaire dans le cadre d'une prédisposition héréditaire au cancer du sein suspectée et qu'aucune mutation constitutionnelle des gènes *BRCA1* et *BRCA2* n'a été identifiée à l'exception de deux patientes, G2014701_III :10 et G2009467_III :11, porteuses respectivement d'une mutation *BRCA1* et *BRCA2*. Les caractéristiques de ces dix familles (type de pathologie, âge au diagnostic et nomenclature des mutations *ATM* et *BRCA*

Famille ID	Npatients	Ntumeurs	Nconstit	Ntumeurs OK	Stratégie Tumorale OK	% Cell. Tumorales	Statut mutationnel BRCA
G1995037	7	3	6	3	oui	60 / ND / ND	wt
G1995042	2	2	2	2	oui	90 / ND	wt
G1995049	6	2	7	1	non	10 / ND	mutation BRCA1
G1995064	2	2	1	2	oui	60 / ND	wt
G1995079	13	2	13	2	oui	55 / 55	wt
G1996019	19	3	18	1	non	≤ 5 / 5 / 60	wt
G2001185	2	2	1	1	non	5-10 / 75	wt
G2002199	2	2	2	2	oui	ND / ND	wt
G2003221	2	2	1	1	non	0 / ND	mutation BRCA1
G2006232	3	3	1	3	oui	60 / 80 / 90	wt
G2007180	2	2	2	1	non	≤ 5 / 5	wt
G2007398	2	2	1	2	oui	25 / 40	wt
G2009159	2	2	2	2	oui	75 / 90	wt
G2009223	2	2	2	1	non	5 / 60	wt
G2009373	2	2	2	2	oui	80 / 85	mutation BRCA2
G2010220	2	2	2	2	oui	20 / 75	wt
G2010325	2	2	1	2	oui	75 / 80	wt
Total	72	37	64	30	11		3

Tableau 2. Caractéristiques des 17 formes familiales de cancer du sein

Famille ID, Identité famille; NPatients, Nombre de patients par famille; Ntumors OK, Nombre d'échantillons tumoraux de bonne qualité par famille; Stratégie Tumorale OK, Stratégie tumorale familiale par SNP-array contributive; % Cell. Tumorales (x / x / x), pourcentage de cellules tumorales dans les échantillons familaux analysés (Tumeur 1 / Tumeur 2 / Tumeur 3); ND, non documenté ; wt, wild-type.
3.2 Méthodes

3.2.1 Extraction d'ADN

3.2.1.1 ADNs tumoraux

La grande majorité des ADNs tumoraux ont été extraits à partir de prélèvements de 50 à 100 mg de chaque tumeur mammaires conservées à -120°C. L'extraction d'ADN a été réalisée dans ce cas selon un protocole classique d'extraction au phénol-chloroforme comprenant une lyse des cellules à la guanidine-HCl, une extraction d'ADN à l'aide de phénol-chloroforme et une précipitation à l'éthanol, ou au moyen d'un kit QIAGEN, le QIAamp® DNA Mini Kit, selon les recommandations du fournisseur.

Certains ADN tumoraux ont été extraits à partir de matériel FFPE (pour *Formalin-fixed, paraffin-embedded*) au moyen d'un kit QIAGEN, le QIAamp® DNA FFPE Tissue kit, selon les recommandations du fournisseur.

Le pourcentage de cellules tumorales a été contrôlé par deux coupes en lame blanche sur le dessus et le dessous du prélèvement.

3.2.1.2 ADNs leucocytaires

Les ADNs leucocytaires ont été extraits à partir de prélèvements sanguins de 5 à 10 ml sur tubes EDTA conservés à 4°C puis -80°C. Pour les prélèvements les plus anciens, l'extraction d'ADN a été réalisée selon un protocole classique d'extraction au phénol-chloroforme comprenant une lyse des cellules à la guanidine-HCl, une extraction d'ADN à l'aide de phénol-chloroforme et une précipitation à l'éthanol. Pour les plus récents, l'extraction a été réalisée sur un automate de type Autopure LS® de QIAGEN, utilisant des réactifs AutoGen, selon le mode opératoire recommandé par le fournisseur.

3.2.2 Génotypage tumoral par puces SNP haute résolution

La réalisation des génotypages tumoraux et leur analyse bioinformatique ont été confiées au programme « Carte d'Identité des Tumeurs » (CIT) de la Ligue Nationale Contre le Cancer, ce projet ayant été sélectionné sous l'appellation KFAM3 par le conseil scientifique Ligue-INCa de ce programme.

Le but de cette approche était d'identifier des régions candidates communes aux tumeurs du sein ayant touché deux à trois membres différents d'une même famille en utilisant des données de puces SNP. Elle impliquait donc (i) la recherche de régions de LOH communes entre les tumeurs familiales, (ii) la reconstruction des haplotypes tumoraux dans ces régions, puis (iii) la recherche d'haplotypes conservés dans les tumeurs étudiées.

3.2.2.1 Hybridation des puces SNP

37 ADNs tumoraux extraits de cancers du sein ont donc été analysés sur des puces SNP HumanOmniExpress v12® d'Illumina, ces dernières contenant chacune 731442 sondes.

L'hybridation des ADNs sur puce a été réalisée par la Société IntegraGen (Evry, France) sous contrat avec le programme CIT selon les instructions fournies par le fournisseur. Les signaux de fluorescence bruts générés ont été importés dans le logiciel BeadStudio® d'Illumina et normalisés comme précédemment décrit à l'aide de mesures de fluorescence effectuées sur des ADNs normaux [378]. Les données de nombre de copies (ou LRR pour *Log R Ratio*) et de statut allélique (ou BAF pour *B allele frequency*) obtenues pour chaque marqueur, sont représentées sur un graphique en fonction de la position chromosomique du SNP.

3.2.2.2 Détection des LOHs

Les aberrations chromosomiques ou les variations du nombre de copies tumorales, ont été déterminées sur la totalité du génome selon la méthode GAP comme précédemment décrit [115]. Les profils génomiques ont été divisés en segments, en appliquant séparément l'algorithme de segmentation binaire circulaire (DNAcopy package, Bioconducteur) aux signaux LRR et BAF [379;380]. Le nombre de copies de l'allèle spécifique de chaque segment a ensuite été déterminé, toujours par la méthode GAP [115]. Brièvement, le modèle GAP a été appliqué pour chaque échantillon et le meilleur modèle obtenu a été utilisé pour déterminer la ploïdie de l'échantillon, le niveau de contamination par des cellules normales ainsi que le nombre de copies absolu et le génotype correspondant à chaque groupe de

segments. Les valeurs du LRR oscillent autour de 0, où LRR = 0 signifie qu'à ce locus, le nombre de copies d'ADN de l'échantillon et le nombre de copies d'ADN du témoin normal sont identiques. Une augmentation du nombre de copies à un locus se traduit par un LRR > 0, et une diminution par un LRR < 0. Nous avons considéré qu'un segment présentait une LOH lorsque le nombre de copies était inférieur à 0,5 et qu'elle-même était significative lorsqu'elle faisait au moins 1 Mb. Cette valeur seuil de 1Mb a été fixée pour éviter d'identifier par la suite des haplotypes identiques par hasard car reposant sur l'étude de peu de SNPs.

3.2.2.3 Reconstruction des haplotypes tumoraux

Dans les régions de LOH, seul un chromosome parental sur les deux est représenté dans la tumeur. L'haplotype du chromosome restant est ainsi directement donné par le génotypage de l'ADN tumoral.

Les haplotypes tumoraux ont donc été reconstruits à partir des profils BAF en rappelant que pour chaque SNP génotypé, deux allèles sont possibles, identifiés de manière arbitraire allèle A et allèle B et que le profil BAF représente le pourcentage d'allèle B à une position génomique donnée. Un rapport de 0,5 traduit ainsi une hétérozygotie A/B parfaite avec 50% d'allèles B et 50% d'allèle A, sans déséquilibre allélique. De même, un rapport de 0 ou 1 traduit une homozygotie parfaite AA ou BB, sans déséquilibre allélique. Dans les tumeurs, on observe fréquemment une déviance de ce rapport, un déséquilibre allélique, car elles sont le siège de nombreux remaniements chromosomiques. Un profil BAF proche de 0 signifie que l'allèle A est majoritairement représenté à ce locus alors qu'un profil BAF proche de 1 signifie que c'est l'allèle B qui est majoritairement représenté. En pratique, des valeurs seuils ont été définies pour attribuer les allèles. Les SNPs avec un BAF >0,65 ont pour génotype l'allèle A. En revanche, les SNPs avec un BAF situé entre ces deux valeurs seuils ont été retirés de l'analyse.

3.2.2.4 Détection des haplotypes conservés

Les régions de LOH communes entre les tumeurs familiales étant établies et les haplotypes tumoraux situés dans ces régions reconstruits, des haplotypes significativement conservés ont pu être recherchés dans chaque région candidate comme précédemment décrit [381]. Brièvement, le pipeline Foudertracker utilisé par Letouzé et al, permet d'assigner un score IBD (pour *Identity By Descent*) à chaque SNP présent dans les régions tumorales étudiées. L'algorithme GERMLINE a d'abord été appliqué pour détecter les segments IBD par paire d'échantillon, cet algorithme tenant compte de la taille et de la fréquence allélique des SNPs contenus dans chaque segment [382]. Le score IBD a ensuite été comparé à une distribution nulle établie à partir d'un set d'haplotypes tumoraux de référence afin d'identifier les SNPs avec des scores IBD significativement plus élevés que ceux attendus par chance. A noter que lorsqu'il y avait plus de deux échantillons tumoraux par famille, l'algorithme Dash était secondairement appliqué [383]. La visualisation de ces données a pu se faire grâce au logiciel IGV® (pour *Integrative Genomics Viewer Broad Institute*, USA).

3.2.3 Séquençage de Nouvelle Génération (NGS) des exomes par synthèse (Illumina) [384;385]

La réalisation des 7 exomes constitutionnels a été confiée à la société IntegraGen.

3.2.3.1 Constitution de la banque de capture [386]

La banque de sondes oligonucléotidiques biotinylées de 120-mers du kit Agilent SureSelect All Exon V5® qui permet de réaliser la capture exomique de 21 522 gènes soit 357 999 exons (50 Mb) à partir d'ADN génomique, a été utilisée. Elle a été construite à partir des bases de données suivantes : CCDS (Consensus Coding Sequences), RefSeq (Reference Sequences), GENCODE (Genome annotation for ENCyclopedia Of DNA Elements), miRBase (microRNA database), TCGA (The Cancer Genome Atlas), UCSC Genome (University of California, Santa Cruz).

Pour chaque exon de chaque gène, une région intronique d'environ 50 bases de part et d'autre de l'exon était ajoutée. La profondeur de capture attendue était d'1X pour 100% des cibles et de 10X pour 94% des cibles.

3.2.3.2 Préparation de la librairie d'ADN génomique

La capture des séquences, l'enrichissement et l'élution ont été effectués selon le protocole et les recommandations du fournisseur (SureSelect, Agilent) sans modification, en vue du séquençage haut débit selon la technologie Illumina.

Brièvement, 3 µg d'ADN génomique, quantifiés au Qubit® (Invitrogen) sont fragmentés par sonication, et purifiés afin d'obtenir des fragments de 150 à 200 pb (paires de bases). Il faut ensuite réparer les extrémités, modifier l'extrémité 3' des fragments (ajout d'une adénine aux extrémités 3'), puis liguer les adaptateurs oligonucléotidiques de séquençage *Paired-End* à chaque extrémité. Ces fragments sont alors de nouveau purifiés et amplifiés par 4 à 6 cycles de PCR. 500ng de ces librairies purifiées sont alors déshydratés avant d'être repris dans un tampon d'hybridation, puis hybridés à la banque de sondes oligonucléotidiques de capture SureSelect All Exon V5® pendant 24h. Après hybridation, lavage et élution, les séquences capturées de chaque échantillon sont de nouveau amplifiées (10 à 12 cycles de PCR) avec ajout d'un index spécifique de chaque échantillon. La fraction obtenue est alors purifiée, qualifiée et quantifiée par PCR quantitative afin d'obtenir suffisamment d'ADN matrice pour la suite du processus en aval.

Le mélange (ou pool) des échantillons est alors réalisé, quantifié et déposé sur l'HiSeq 2000® d'Illumina pour séquençage, par synthèse de séquences de 75 bases de chaque extrémité.

3.2.3.3 Séquençage sur HiSeq 2000® (Illumina)

C'est une technologie de séquençage sur puce. Chaque cycle comprend 3 étapes : Incorporation, détection et déprotection. Cette technique est fondée sur l'intégration de plusieurs méthodes :

- o Les biopuces à ADN (ou flowcell),
- Une variante de la technique de Sanger appelé CRT (pour *cyclic reversible termination*),
- Un traitement informatique pour l'acquisition, le traitement et l'analyse des images.
 - *Flowcell* (ici PE Flow Cell V3 8 échantillons/flowcell)

La très haute densité de la puce de séquençage ou flowcell (plus de 100 millions de molécules par centimètre carré) permet de séquencer environ 100000 paires de bases par seconde. L'ADN est coulé dans des cellules spécialement conçues pour assurer la fixation de chaque molécule d'ADN obtenue, après amplification en phase solide de la première molécule d'ADN fixé grâce aux adaptateurs. La formation des ponts d'amplification permet d'avoir une haute densité de brins d'ADN ce qui facilite et augmente le débit de séquençage.

CRT (pour cyclic reversible termination)

Le principe de ce séquençage est basé sur l'incorporation réversible de nucléotides fluorescents et la lecture optique de cette fluorescence. Comme pour la technique de Sanger, il s'agit d'une terminaison de synthèse fondée sur l'utilisation d'un terminateur (mais celui-ci est réversible) contenant un groupement de protection attaché au nucléotide qui termine la synthèse d'ADN. L'élimination du groupement de protection par photoclivage utilisant la lumière ultraviolette (>300nm) permet la restauration du groupement fonctionnel du nucléotide incorporé, ce qui permet à l'ADN polymérase d'incorporer le prochain nucléotide et ainsi de suite. Les molécules de 300 paires de bases environ sont séquencées à chaque extrémité (ici 75 bases) puis alignées sur la séquence du génome humain de référence pour être reconstituée.

3.2.3.4 Analyses bioinformatiques primaire, secondaire et tertiaire des données NGS

L'analyse bio-informatique primaire consiste en la conversion en temps réel des 112 images capturées à chaque cycle d'incorporation de nucléotide (4 fois 28 *tiles*) : pour chaque cycle, il y a 4 images capturées correspondant à chacun des 4 nucléotides potentiellement incorporés, réalisées pour chacune des 28 zones (*tile*) de la *flowcell*. Le séquençage réalisé est un séquençage paired-end à une profondeur moyenne de 60X, c'est-à-dire que chaque base est lue en moyenne 60 fois. Le séquenceur traduit ensuite ces signaux en séquences (*« basecalling »*) et un score de qualité est assigné à chaque base. Les logiciels gérant cette phase de NGS sont HiSeq Control Software 2.2.58® (HCS) et Real Time Analysis 1.18.64.0® (RTA).

L'analyse secondaire et la première partie de l'analyse tertiaire se résument en deux grandes étapes : l'alignement des séquences sur le génome et la détection des variants. On regroupe sous le terme de « *pipeline* » la succession des différents algorithmes et programmes utilisés pour réaliser l'analyse. Ici, l'analyse des données de séquençage fournie par IntegraGen utilise le pipeline fourni par Illumina, CASAVA 1.8.2® qui comprend l'alignement des séquences sur le génome humain de référence hg19, et une détection des variants SNVs (pour *single nucleotide variation*) et INDELs (pour *Insertions-Délétions*).

Les fichiers bruts de format .bcl issus de l'analyse primaire, ont été d'abord convertis en fichiers de format .fastq qui permettent de stocker les séquences nucléiques et leurs scores de qualité. Le score de qualité utilisé est le QPhred, qui correspond à une probabilité p d'erreur de mauvaise identification de la base, codée en -10 $\log_{10}p$. Ainsi, un QPhred de 30 correspond à une probabilité d'erreur de 0,1%. Puis, à partir des données .fastq, les séquences ou reads ont été alignées sur le génome de référence grâce à l'algorithme ELANDv2e®, et ont été converties en fichiers de format .sam (Sequence Alignment/Map). Les fichiers .sam ont été ensuite compressés, indexés et convertis en fichiers .bam, qui permettent de visualiser les reads alignés, sous IGV® (Integrative Genomics Viewer Broad Institute, USA) ou Alamut® (Interactive Biosoftware, Rouen, France). Les reads dupliqués, ou avec une taille d'alignement trop courte, ou avec une qualité d'alignement trop faible ont été éliminés. Cette étape a été réalisée par le module « sort » de CASAVA. Les fichiers .bam ont subi alors un nouveau réalignement car les INDELs peuvent créer de faux SNVs, et également une recalibration des scores de qualité des bases par le module « CallSmallVariants » de CASAVA. A partir de ces données, des fichiers de format .vcf (variant call format) ont été générés au niveau des régions cibles. Ils permettent de visualiser les différentes variations détectées, simultanément aux fichiers .bam.

Les fichiers .vcf bruts ne sont pas annotés, cela signifie que les variants rapportés sont exprimés en position génomique, allèle de référence/allèle variant et données de qualité comme la profondeur de lecture et le score de qualité du variant. Ici, l'annotation a été réalisée en interne par Integragen incluant :

- Une annotation des gènes par les bases de données Refseq et Ensembl.
- Une annotation des variants référencés par les bases de données dbSNP version 132, 1000Genomes version V3-20101123, HapMap version 27, et EVS (*« Exome Variant*

Server ») version ESP6500SI-V2-SSA137, qui donnent des indications sur la fréquence allélique en population générale des variants.

- Une annotation des variants référencés par la base de données d'Integragen, qui détermine des fréquences exomiques (HMZ (pour homozygote) / HTZ (pour hétérozygote) sur la base de 201 exomes pour les SNVs, et 130 exomes pour les INDELs.
- Des caractérisations des variants : intronique/exonique/UTR, silencieuse/fauxsens/non-sens/au niveau des sites consensus d'épissage, de statut homozygote/ hétérozygote non basé sur le ratio allélique mais sur un algorithme complexe de probabilité de statut.
- Une annotation des SNVs par l'algorithme de prédiction Polyphen.
- Des informations sur les conséquences des variants situés sur la séquence transcrite et codante : position du variant dans la séquence codante, position du variant dans la séquence protéique, description du codon sauvage/muté, descriptif de l'acide aminé sauvage/muté.

A l'issue de l'analyse CASAVA 1.8, pour chaque échantillon, un fichier Excel recensant les variants de type SNV et un fichier recensant les variants de type INDEL ont été générés.

3.2.4 Séquençage de Nouvelle Génération (NGS) ciblé par synthèse (Illumina) [384;385]

14 ADNs tumoraux (de cancer du sein) ont bénéficié d'une analyse NGS ciblé sur un panel de gènes. Cette analyse a été réalisée dans le laboratoire d'Oncogénétique de l'Institut Bergonié.

3.2.4.1 Constitution de la banque de Capture

Une banque d'oligonucléotides de 120-mers a été utilisée pour réaliser la capture exomique de 25 gènes impliqués essentiellement dans la prédisposition héréditaire aux cancers du sein et/ou des ovaires, mais également dans la prédisposition héréditaire aux cancers digestifs et autres syndromes plus rares et plus complexes comme la maladie de Cowden et le syndrome de Gorlin. La liste des gènes de la banque de capture NGS avec leurs principales

caractéristiques (localisation chromosomique, coordonnées génomiques, orientation du gène, nombre d'exons, taille du transcrit, identifiants des gènes selon UCSC, RefSeq NCBI et Ensembl) est fournie dans l'article (Sup Table 1).

La banque de capture a été construite à l'aide du logiciel de design Suredesign de Agilent technologies (<u>https://earray.chem.agilent.com/suredesign/</u>). La profondeur de capture est réalisée au niveau 2X. Les coordonnées génomiques des 451 exons des 25 gènes ont été obtenues par soumission des identifiants UCSC au site BioMart Tools d'Ensembl (<u>http://www.ensembl.org/index.html</u>). Pour chaque exon de chaque gène, une région intronique de 50 bases de part et d'autre de l'exon a été ajoutée. Les coordonnées génomiques ainsi définies ont été soumises au design (Suredesign) avec comme critère de design :

- Profondeur de capture 2X
- Repeat Masker : stringent
- Boosting : Maximize performance

Une banque de 3000 sondes oligonucléotidiques uniques a été définie, répliquées de 19 à 38 fois selon le pourcentage de GC dans les régions capturées. Les régions génomiques exoniques mal couvertes ont été reprises une à une de façon à forcer le logiciel à définir une sonde oligonucléotidique de capture dans ces régions. Les 316 sondes ainsi définies ont été soumises à *Blat* (<u>http://genome.ucsc.edu/cgi-bin/hgBlat?command=start</u>) par batch de 25 afin de s'assurer d'une couverture « *off-target* » minimale.

3.2.4.2 Préparation de la librairie d'ADN tumoraux

La capture des séquences, l'enrichissement et l'élution ont été effectués selon le protocole et les recommandations du fournisseur (kit SureSelect XT, Agilent) sans modification, en vue du séquençage haut débit selon la technologie Illumina.

Brièvement, les échantillons sont quantifiés au Qubit (Invitrogen) et qualifiés sur TapeStation 2200 (Agilent) pour évaluer l'intégrité du matériel initial sous la forme d'une valeur DIN (pour *DNA Integrity Number*), puis 200ng d'ADN tumoral sont fragmentés par sonication (instrument Covaris) afin de sélectionner des fragments de 150 à 200 pb. Il faut ensuite réparer les extrémités, modifier l'extrémité 3' des fragments (ajout d'une adénine aux extrémités 3'), puis liguer les adaptateurs oligonucléotidiques de séquençage *Paired-End* à

chaque extrémité. Ces fragments sont alors amplifiés par 10 cycles de PCR, puis qualifiés et quantifiés sur TapeStation avec une taille attendue entre 225 et 275 pb. 500ng de ces librairies purifiées sont alors déshydratés, puis hybridés à la banque de sondes oligonucléotidiques de capture pendant 16 heures. Après hybridation, capture et lavage, les séquences capturées de chaque échantillon sont de nouveau amplifiées (16 cycles de PCR) avec ajout d'un index spécifique de chaque échantillon. La fraction obtenue est alors purifiée puis qualifiée et quantifiée sur Bioanalyzer 2100 (Agilent) avec une taille de fragment attendue entre 250 et 350 pb.

Le mélange (ou pool) des échantillons est alors réalisé, quantifié et déposé sur le MiSeq d'Illumina pour séquençage par synthèse.

3.2.4.3 Séquençage sur MiSeq (Illumina)

Le séquençage sur MiSeq, comme précédemment sur HiSeq, utilise une technologie de séquençage sur puce. Il comprend les mêmes étapes et est fondé sur l'intégration des mêmes méthodes (cf paragraphe 3.2.3.3).

3.2.4.4 Analyses bioinformatiques primaires, secondaire et tertiaires des données NGS

L'analyse bio-informatique primaire est similaire à celle précédemment décrite pour l'analyse des exomes, le logiciel gérant cette phase de NGS étant Real Time Analysis 1.18.64.0® (RTA).

L'analyse secondaire, et la première partie de l'analyse tertiaire se résument comme précédemment en deux grandes étapes : l'alignement des séquences sur le génome et la détection des variants. Le « *pipeline* » utilisé est cependant un peu différent.

Les fichiers bruts de format .bcl issus de l'analyse primaire, ont été d'abord convertis en fichiers de format .fastq par le logiciel MSR [387]. A partir de ces données, des fichiers de format .sam (*Sequence Alignment/Map*) ont été générés grâce à l'algorithme BWA [388] permettant l'alignement des séquences ou reads au génome de référence hg19. Les fichiers .sam ont été ensuite compressés, indexés et convertis en fichiers .bam, permettant de visualiser les *reads* alignés sous IGV® ou Alamut®. De la même manière que pour les

analyses d'exome, les *reads* dupliqués, ou avec une taille d'alignement trop courte, ou avec une qualité d'alignement trop faible ont été éliminés mais grâce à un outil différent, Picard® (v.1.119, <u>http://broadinstitute.github.io/picard/</u>), et les fichiers .bam ont subi un nouveau réalignement ainsi qu'une recalibration des scores de qualité des bases à l'aide de l'outil GATK [389]. Des variants somatiques potentiels, de type INDEL et SNV, ont pu être remontés grâce à l'outil Somatic Variant Caller® développé par Illumina et annotés par Ensembl VEP [390]. A partir de ces données, des fichiers de format .vcf (*variant call format*) ont été générés au niveau des régions cibles permettant de visualiser les différentes variations détectées, simultanément aux fichiers .bam.

Pour cette analyse nous avons uniquement considéré les variants passant les filtres standards fixés par Somatic Variant Caller® et concernant les gènes *ATM*, *BRCA1* et *BRCA2*. A l'issue de cette analyse, pour chaque échantillon, un fichier Excel recensant les variants détectés a été généré. En moyenne, la profondeur par base était de 630X (74X - 1134X) et 99% des régions cibles étaient couvertes à 378X minimum.

3.2.5 Séquençage Sanger

Les mutations constitutionnelles du gène *ATM*, détectées par NGS chez les cas index, ont toutes été confirmées par séquençage conventionnel selon la méthode de Sanger [181;182]. Cette méthode de séquençage a également été utilisée pour étudier la ségrégation de ces mutations avec les atteintes mammaires chez les apparentés disponibles.

3.2.5.1 Amplification par réaction de polymérisation en chaîne ou PCR (pour *Polymerase Chain Reaction*)

Brièvement, le programme informatique Primer3 (http://frodo.wi.mit.edu/) a été utilisé pour sélectionner les couples d'amorces au niveau de chaque exon à amplifier. Puis le programme BLAST (http://www.ncbi.nlm.nih.gov.gate2.inist.fr/blast/Blast.cgi) a été utilisé pour effectuer un alignement de séquence pour chacune des amorces de façon à vérifier l'absence d'homologie avec des séquences hétérologues et garantir la spécificité des PCR. L'ADN leucocytaire ou tumoral a ensuite été amplifié par PCR en utilisant le kit AmpliTaq Gold® (Life Technologies, Paisley, UK) selon les recommandations du fournisseur. Les couples d'amorces, le mix réactionnel ainsi que le programme de PCR utilisés pour le séquençage sont

respectivement présentés en annexe (Annexes 18, 19 et 20). Un contrôle électrophorétique sur gel d'agarose 2% a été réalisé avant de purifier 5µl des produits de PCR obtenus à l'aide du kit ExoSAP-IT® (Affymetrix).

3.2.5.2 Réactions de séquence

Les réactions de séquence ont été réalisées grâce au kit BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) sur un thermocycleur SimpliAmp (Life Technologies). Le mix réactionnel est présenté en annexe (Annexe 21) et le programme utilisé pour les réactions de séquence était le suivant : 10 sec à 96°C, 5 sec à 50°C et 4 min à 60°C, le tout pendant 25 cycles.

Les produits de séquence ont ensuite été purifiés sur billes magnétiques à l'aide du kit de purification BigDye XTerminator® (Applied Biosystems) selon les recommandations du fournisseur. Enfin, les produits de purification ont été analysés sur séquenceur capillaire ABI 3130x1 (Applied Biosystems). Les électrophorégrammes résultants ont été analysés à l'aide du logiciel Sequence Analysis (Applied Biosystems) puis alignés et comparés à la séquence génomique de référence à l'aide du logiciel Seqscape (Applied Biosystems).

4 RÉSULTATS

4.1 Etude 1 : A propos de 17 formes familiales de cancer du sein

4.1.1 Recherche de régions candidates par Génotypage tumoral

Des études sur puces SNP ont donc été menées en collaboration avec le programme Carte d'Identité des Tumeurs sur 37 tumeurs du sein provenant de 17 familles distinctes. On rappelle que cette approche visait à identifier des régions candidates de taille restreinte permettant de guider l'analyse des séquences obtenues par séquençage NGS. Ceci impliquait (i) la recherche de régions de LOH communes entre les tumeurs du sein familiales explorées et (ii) la recherche d'haplotypes conservés au sein de ces dernières.

4.1.1.1 Echec d'analyse

Un contrôle qualité des profils SNP obtenus pour chaque tumeur analysée, a montré que la stratégie initialement prévue, nécessitant au moins l'analyse de deux tumeurs par famille, ne pouvait être appliquée qu'à 11 familles sur 17 en raison d'une contamination importante de 7/37 échantillons tumoraux par des cellules non tumorales. Les profils SNP de ces échantillons étaient effectivement plats, peu remaniés et sans déséquilibre associé, témoignant de l'analyse de cellules normales plus que tumorales (Figures 6 à 11). L'évaluation histologique de la proportion de cellules tumorales des échantillons tumoraux était d'ailleurs en accord avec les données de puces SNP en affichant pour ces derniers, un pourcentage de cellules tumorales souvent très inférieur à 10% (Tableau 2). Six familles (G1995049, G1996019, G2001185, G2003221, G2007180 et G2009223) dont les deux familles témoins d'une mutation *BRCA1* (G1995049 et G2003221) et une famille très évocatrice d'une prédisposition héréditaire au cancer du sein car neuf cas autour de l'âge de 40 ans y étaient rapportés (famille G1996019), ont donc été éliminées de notre étude.



Figure 6: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995049.

Les tumeurs HS_160 et HS_162 de la famille G1995049 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Le profil génomique de la tumeur HS_162 ne présentant pas, ou pratiquement pas, de remaniement et de déséquilibre allélique, montre que l'échantillon analysé est contaminé par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.



Figure 7: Profils génomiques obtenus par puces SNP pour les trois cancers du sein de la famille G1996019. Les tumeurs HS_191, HS_193 et HS_195 de la famille G1996019 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Les profils génomiques des tumeurs HS_193 et HS_195 ne présentant pas, ou pratiquement pas, de remaniement et de déséquilibre allélique, montrent que les échantillons analysés sont contaminés par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.



Figure 8: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2001185.

Les tumeurs HS_212 et HS_213 de la famille G2001185 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Le profil génomique de la tumeur HS_213 présentant un bruit de fond important du LRR et pratiquement pas de déséquilibre allélique associé aux remaniements observés, montre que l'échantillon analysé est contaminé par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.



Figure 9: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2003221.

Les tumeurs HS_223 et HS_224 de la famille G2003221 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Le profil génomique de la tumeur HS_223 ne présentant pas, ou pratiquement pas, de remaniement et de déséquilibre allélique, montre que l'échantillon analysé est contaminé par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.



Figure 10: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2007180.

Les tumeurs HS_230 et HS_232 de la famille G2007180 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Le profil génomique de la tumeur HS_230 mais également celui de la tumeur HS_232 ne présentant pas, ou pratiquement pas, de remaniement et de déséquilibre allélique, montrent que les échantillons analysés sont contaminés par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.



Figure 11: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009223. Les tumeurs HS_241 et HS_243 de la famille G2009223 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Le profil génomique de la tumeur HS_243 ne présentant pas, ou pratiquement pas, de remaniement et de déséquilibre allélique, montrent que l'échantillon analysé est contaminé par un excès de cellules non tumorales ne permettant la poursuite des investigations par paire ou trio de cancers du sein familiaux.

4.1.1.2 Détection de LOHs communes dans les tumeurs familiales

Les données de puces SNP ont pu être analysées par paire ou trio de tumeurs du sein pour 11 familles.

La 1^{ère} étape de l'analyse consistant à rechercher des régions de LOH communes significatives (>1Mb) entre les tumeurs familiales, a montré qu'aucune n'était identifiée pour trois familles : les familles G1995064, G2007398 et G2010220 (Figures 12 à 15). Concernant la famille G2009373, témoin d'une mutation *BRCA2*, deux régions de LOH communes aux deux tumeurs familiales ont été identifiées, sur les chromosomes 6 [6q15 - 6q27] et 17 [17p13.3 - 17p12], mais aucune au locus du gène *BRCA2* [13q12-13]. Sur la Figure 15, on observe effectivement que la tumeur du sein de la grand-mère maternelle (patiente IV:1) ne présente pas de LOH en regard du chromosome 13 alors qu'elle est observée pour la tumeur de sa petite-fille (patiente II:12). Cet exemple montre que le gène suppresseur de tumeur *BRCA2*, n'est pas obligatoirement la cible d'une LOH dans la tumeur, pouvant être inactivé par un autre moyen tel qu'une mutation ou un mécanisme épigénétique, non recherché ici.

En revanche, des régions de LOH communes ont été mises en évidence pour les tumeurs familiales de sept familles (Figures 16 à 22). Au total, 45 régions de LOH ont été détectées, soit : deux pour la famille G1995037, cinq pour la famille G1995042, trois pour la famille G1995079, trois pour la famille G2002199, 16 pour la famille G2006232, 11 pour la famille G2009159 et cinq pour la famille G2010325. Les caractéristiques de ces différentes régions, en termes de taille et de localisation exactes, sont détaillées par famille dans le Tableau 3.

4.1.1.3 Détection d'haplotypes conservés dans les régions de LOH

L'analyse des régions de LOH précédemment identifiées a permis ensuite de détecter des haplotypes conservés pour les sept familles.

Un total de 74 haplotypes candidats a ainsi été identifié. Les caractéristiques de ces derniers (leur localisation précise, leur taille et le nombre de gènes qu'ils contiennent) sont présentées par famille dans les Tableaux 4 à 10.



Figure 12: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995064 et absence de mise en évidence de région de LOH commune supérieure à 1 Mb.

Les tumeurs HS_173 et HS_174 de la famille G1995064 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, aucune région de LOH commune n'a été mise en évidence pour les deux tumeurs familiales étudiées, les quelques LOH(s) détectées étant inférieures à 1 Mb.





Les tumeurs HS_234 et HS_235 de la famille G2007398 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, aucune région de LOH commune n'a été mise en évidence pour les deux tumeurs familiales étudiées, les quelques LOH(s) détectées étant inférieures à 1 Mb.





Les tumeurs HS_215 et HS_217 de la famille G2010220 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, aucune région de LOH commune n'a été mise en évidence pour les deux tumeurs familiales étudiées, les quelques LOH(s) détectées étant inférieures à 1 Mb.



Figure 15: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009373 (Témoin *BRCA2*) et absence de mise en évidence de région de LOH commune au locus du gène *BRCA2* (13q12-13).

Les tumeurs HS_245 et HS_247 de la famille G2009373 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). A noter que les patientes ayant présenté ces tumeurs sont porteuses d'une mutation constitutionnelle délétère du gène *BRCA2*. (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, deux régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales étudiées, sur les chromosomes 6 et 17. En revanche, la tumeur HS_245 ne présente pas de LOH au locus 13q12-13 du gène *BRCA2* (représenté ici par un cadre noir) alors que la tumeur HS_247 en présente bien une.



Figure 16: Profils génomiques obtenus par puces SNP pour les trois cancers du sein (lié 2 à 2) de la famille G1995037 et mise en évidence de régions de LOH communes.

Les tumeurs HS_151, HS_153 et HS_154 de la famille G1995037 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). A noter que ces trois tumeurs sont en fait liées généalogiquement deux à deux (HS_151 avec HS_154 et HS_153 avec HS_154). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, deux régions de LOH(s) communes ont été identifiées pour les trois tumeurs familiales étudiées, sur les chromosomes 11 et 16. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.



Figure 17: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995042 et mise en évidence de régions de LOH communes.

Les tumeurs HS_169 et HS_171 de la famille G1995042 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, cinq régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales étudiées, sur les chromosomes 1, 6 (2 régions), 9 et 14. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.



Figure 18: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G1995079 et mise en évidence de régions de LOH communes.

Les tumeurs HS_176 et HS_178 de la famille G1995079 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, trois régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales étudiées, sur les chromosomes 6, 11 et 22. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.



Figure 19: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2002199 et mise en évidence de régions de LOH communes.

Les tumeurs HS_219 et HS_221 de la famille G2002199 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, trois régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales étudiées, sur les chromosomes 11, 13 et 17. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.





Les tumeurs HS_226, HS_227 et HS_228 de la famille G2006232 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, seize régions de LOH(s) communes ont été identifiées pour les trois tumeurs familiales étudiées, sur les chromosomes 1 (3 régions), 6, 8, 11 (2 régions), 12, 13 (4 régions), 17, 18 (2 régions) et 22. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.



Figure 21: Profils génomiques obtenus par puces SNP pour les deux cancers du sein de la famille G2009159 et mise en évidence de régions de LOH communes.

Les tumeurs HS_237 et HS_239 de la famille G2009159 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb. Dans ces conditions, onze régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales

étudiées, sur les chromosomes 4, 6, 8, 9, 10, 13, 15, 17, 19 (2 régions) et 22. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.





Les tumeurs HS_249 et HS_250 de la famille G2010325 ont été analysées par puce SNP HumanOmniExpress v12 (Illumina). (A) Le nombre de copies spécifique de l'allèle (ou *allele-specific copy number*) de chaque segment a été déterminé par la méthode GAP (pour *Genome Alteration Print*) et il est représenté par le nombre de copies absolu. (B) La variation du nombre de copies est représentée par le log R ratio (ou LRR), centré sur zéro. (C) Les déséquilibres alléliques sont représentés par la fréquence de l'allèle B ou BAF (pour *B-allele frequency*). (D) Représentation de la fréquence des LOH(s) observées dans les échantillons tumoraux explorés. Nous avons considéré qu'un segment présentait une LOH quand le nombre de copies était inférieur à 0,5 et que cette LOH était significative quand elle était supérieure à 1 Mb.

Dans ces conditions, cinq régions de LOH(s) communes ont été identifiées pour les deux tumeurs familiales étudiées, sur les chromosomes 3, 9, 13, 17 et 21. Les quelques autres LOH(s) communes détectées étaient non significatives car inférieures à 1 Mb.

Famille	LOH_ID	Chrom	Startpos	Endpos	Startcyto	Endcyto	Length.bp
C1005027	G1995037_11	11	78928731	132815368	11q14.1	11q25	53886637
1992037	G1995037_16	16	57719586	87802311	16q21	16q24.2	30082725
	G1995042_1	1	97999687	98292937	1p21.3	1p21.3	293250
	G1995042_6.1	6	203878	10231925	6p25.3	6p24.3	10028047
G1995042	G1995042_6.2	6	65125567	170919470	6q12	6q27	105793903
	G1995042_9	9	86920632	125044199	9q21.33	9q33.2	38123567
	G1995042_14	14	39802186	107287663	14q21.1	14q32.33	67485477
	G1995079_6	6	61891118	170919470	6q11.1	6q27	109028352
G1995079	G1995079_11	11	90070103	134934063	11q14.3	11q25	44863960
	G1995079_22	22	17177470	47310536	22q11.1	22q13.31	30133066
	G2002199_11	11	71626089	134934063	11q13.4	11q25	63307974
G2002199	G2002199_13	13	43057549	115103529	13q14.11	13q34	72045980
	G2002199_17	17	8547	22242355	17p13.3	17p11.1	22233808
	G2006232_1.1	1	1425700	25039867	1p36.33	1p36.11	23614167
	G2006232_1.2	1	74712181	80248531	1p31.1	1p31.1	5536350
	G2006232_1.3	1	112783743	114861832	1p13.2	1p13.2	2078089
	G2006232_6	6	145122139	162627290	6q24.2	6q26	17505151
	G2006232_8	8	164984	36628405	8p23.3	8p11.23	36463421
	G2006232_11.1	11	198510	12122574	11p15.5	11p15.3	11924064
	G2006232_11.2	11	72800970	134934063	11q13.4	11q25	62133093
62006222	G2006232_12	12	191619	18276749	12p13.33	12p12.3	18085130
92000232	G2006232_13.1	13	63110606	64733435	13q21.31	13q21.31	1622829
	G2006232_13.2	13	76310850	77488736	13q22.2	13q22.3	1177886
	G2006232_13.3	13	86296819	87860052	13q31.1	13q31.2	1563233
	G2006232_13.4	13	98203727	111180101	13q32.2	13q34	12976374
	G2006232_17	17	8547	13946187	17p13.3	17p12	13937640
	G2006232_18.1	18	14308620	15375878	18p11.21	18p11.21	1067258
	G2006232_18.2	18	21178845	41191319	18q11.2	18q12.3	20012474
	G2006232_22	22	32930263	51211392	22q12.3	22q13.33	18281129

LOH, Loss Of Heterozygosity; LOH_ID, Identité de la région de LOH ; Chrom, Chromosome; Startpos/Startcyto, Début de la position génomique et cytologique des LOHs identifiées; Endpos/Endcyto, Fin de la position génomique et cytologique des LOHs identifiées; Nsnp, Length.bp, taille de la LOH en paire de bases.

Famille	LOH_ID	Chrom	Startpos	Endpos	Startcyto	Endcyto	Length.bp
	G2009159_4	4	71566	190915650	4p16.3	4q35.2	190844084
	G2009159_6	6	67860579	170353312	6q12	6q27	102492733
	G2009159_8	8	164984	41877317	8p23.3	8p11.21	41712333
	G2009159_9	9	130598125	141066491	9q34.11	9q34.3	10468366
	G2009159_10	10	61331409	135477883	10q21.2	10q26.3	74146474
G2009159	G2009159_13	13	19058717	115103529	13q11	13q34	96044812
	G2009159_15	15	43101885	71993353	15q15.2	15q23	28891468
	G2009159_17	17	8547	47448346	17p13.3	17q21.33	47439799
	G2009159_19.1	19	27795706	33289811	19q11	19q13.11	5494105
	G2009159_19.2	19	52654833	59097160	19q13.41	19q13.43	6442327
	G2009159_22	22	16114244	51211392	22q11.1	22q13.33	35097148
	G2010325_3	3	843748	83952437	3p26.3	3p12.1	83108689
	G2010325_9	9	46587	44218056	9p24.3	9p11.2	44171469
G2010325	G2010325_13	13	19058717	44574009	13q11	13q14.11	25515292
	G2010325_17	17	8547	18923818	17p13.3	17p11.2	18915271
	G2010325_21	21	10827533	48100155	21p11.2	21q22.3	37272622

Tableau 3. Caractéristiques des 45 régions de LOH identifiées (suite)

LOH, Loss Of Heterozygosity; LOH_ID, Identité de la région de LOH ; Chrom, Chromosome; Startpos/Startcyto, Début de la position génomique et cytologique des LOHs identifiées; Nsnp, Length.bp, taille de la LOH en paire de bases.

Tableau 4. Ca	ractéristiques	des 7 ha _l	olotypes co	nservés dé	tectés pour	les trois	cancers du se	ein* analys	és dan:	s la famille G	1995037
	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G1995037_11	HS_151,HS_154	11	91320978	100633303	11q14.3	11q22.1	rs11019558	rs7126324	2762	9312325	43
G1995037_11	HS_151,HS_154	11	101594826	107841291	11q22.1	11q22.3	rs12360583	rs4028250	1515	6246465	39
G1995037_11	HS_153,HS_154	11	120379057	120644424	11q23.3	11q23.3	rs2247168	rs948027	110	265367	1
G1995037_11	HS_151,HS_154	11	120379057	131077439	11q23.3	11q25	rs2247168	rs1793562	3466	10698382	98
G1995037_16	HS_153,HS_154	1 16	70779144	71350049	16q22.1	16q22.2	rs2305688	rs6499502	61	570905	ŝ
G1995037_16	HS_153,HS_154	1 16	75840629	76905308	16q23.1	16q23.1	rs17765038	rs9924808	257	1064679	1
G1995037_16	HS_153,HS_154	1 16	81989484	82451159	16q23.3	16q23.3	rs17793122	rs1025065	175	461675	4
LOH, Loss Of Het	erozygosity; LOH_I	D, Identité	de la région d	e LOH ; Tsamp	oles, Echantill	on tumoral;	Chrom, Chromo	some; Startpo	os/Starto	:yto/StartSNP, D	ébut de la
position génomi	que/cytologique/S	NP des hap	olotypes conse	ervés; Endpos	:/Endcyto/End	SNP, Fin de	la position génc	omique/cytolo	ogique/SI	NP des haploty	oes
conservés; Nsnp	, Nombre de SNPs (contenus d	ans chaque ha	aplotype; Leng	gth.bp, taille d	e l'haplotyp	oe en paire de ba	ases; Ngenes,	Nombre	de gènes contei	ius dans
chaque haploty;	oe; * A noter que ce	s trois can	cers ne sont p	as liés entre-	eux mais deux	: à deux car	appartenant à (des lignées pa	arentales	différentes.	

Tableau 5. Ca	ractéristiques d	es 3 hap	lotypes col	nservés dét	tectés poul	r les deux	cancers du :	sein analys	és dans	la famille G	1995042
п_нол	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G1995042_14	HS_169,HS_171	14	39821153	47633066	14q21.1	14q21.3	rs10130181	rs7140778	1502	7811913	13
G1995042_14	HS_169,HS_171	14	54186083	55583424	14q22.2	14q22.3	rs169748	rs8009181	393	1397341	10
G1995042_14	HS_169,HS_171	14	101356010	107281980	14q32.2	14q32.33	rs3809404	rs1544317	1072	5925970	141
LOH, Loss Of Het	erozygosity; LOH_ID	, Identité (de la région de	e LOH ; Tsamp	oles, Echantill	on tumoral;	Chrom, Chrom	iosome; Startp	os/Starto	cyto/StartSNP, I	Début de
la position géno	mique/cytologique/	SNP des h	aplotypes cor	iservés; Endp	os/Endcyto/E	indSNP, Fin o	le la position g	génomique/cyt	tologique	e/SNP des haple	otypes
conservés; Nsnp,	o, Nombre de SNPs co	ontenus da	ans chaque ha	plotype; Leng	th.bp, taille d	ie l'haplotyp	e en paire de b	oases; Ngenes,	Nombre	de gènes conte	nus dans
chaque haplotyp	De.										

Tableau 6. Ca	iractéristiques o	des 7 hap	olotypes co	nservés dé	tectés poul	r les deux	cancers du :	sein analyse	és dans	la famille G	1995079
	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G1995079_6	HS_176,HS_178	9	61891118	66164344	6q11.1	6q12	rs4502981	rs12206730	566	4273226	5
G1995079_6	HS_176,HS_178	9	90739172	170919470	6q15	6q27	rs3857498	rs12530134	20136	80180298	363
G1995079_11	HS_176,HS_178	11	90070103	131888483	11q14.3	11q25	rs1507608	rs11525415	12139	41818380	315
G1995079_22	HS_176,HS_178	22	22710758	23258369	22q11.22	22q11.22	rs17578409	rs451115	53	547611	7
G1995079_22	HS_176,HS_178	22	34799261	35964722	22q12.3	22q12.3	rs5755023	rs5750116	373	1165461	9
G1995079_22	HS_176,HS_178	22	39575848	41415644	22q13.1	22q13.2	rs12166894	rs4821981	319	1839796	26
G1995079_22	HS_176,HS_178	22	44794360	47310536	22q13.31	22q13.31	rs6007363	rs4823894	951	2516176	34
LOH, Loss Of Het	erozygosity; LOH_II	D, Identité	de la région d	e LOH ; Tsamp	oles, Echantill	lon tumoral;	Chrom, Chrom	osome; Startp	os/Starte	cyto/StartSNP,	Début de
la position géno	mique/cytologique	/SNP des h	aplotypes cor	nservés; Endp	ios/Endcyto/E	EndSNP, Fin o	le la position g	génomique/cyt	tologique	<pre>s/SNP des hapl</pre>	otypes
conservés; Nsnp	, Nombre de SNPs c	contenus d	ans chaque ha	aplotype; Leng	gth.bp, taille c	de l'haplotyp	e en paire de b	ases; Ngenes,	Nombre	de gènes conte	inus dans
chaque haploty	De.										

Tableau 7. C	aractéristiques d	les 4 hap	lotypes co	nservés dét	tectés poul	r les deux	cancers du	sein analys	és dans	la famille G	2002199
	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G2002199_11	HS_219,HS_221	11	71626089	76301375	11q13.4	11q13.5	rs7113974	rs7927997	1021	4675286	61
G2002199_11	HS_219,HS_221	11	77855357	134381569	11q14.1	11q25	rs11237414	rs4606492	16329	56526212	371
G2002199_13	HS_219,HS_221	13	104352252	104941133	13q33.1	13q33.2	rs4145275	rs7334334	193	588831	0
G2002199_17	HS_219,HS_221	17	13043	22213908	17p13.3	17p11.1	rs17054921	rs12325748	6501	22200865	369
LOH, Loss Of He	terozygosity; LOH_ID), Identité	de la région d	e LOH ; Tsamp	les, Echantill	on tumoral;	Chrom, Chrom	iosome; Startp	os/Starte	cyto/StartSNP, [Jébut de
la position géne	omique/cytologique/	/SNP des h	aplotypes cor	iservés; Endp	os/Endcyto/E	indSNP, Fin o	le la position g	génomique/cyt	tologique	e/SNP des haplo	types
conservés; Nsnj	p, Nombre de SNPs co	ontenus d	ans chaque ha	aplotype; Leng	th.bp, taille c	le l'haplotyp	e en paire de k	oases; Ngenes,	Nombre	de gènes conte	nus dans
chaque haploty	pe.										

Tableau 8. Cai	ractéristiques des 19 h	aplotyp	es conserv	és détectés	pour les t	rois cancer	s du sein ar	alysés dan:	s la fan	nille G20062	32
п_нол	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G2006232_1.1	HS_226,HS_227	1	5709833	6100313	1p36.31	1p36.31	rs709184	rs1295087	120	390480	2
G2006232_1.1	HS_226,HS_228	1	5758772	6141325	1p36.31	1p36.31	rs7521197	rs3789541	108	382553	2
G2006232_1.1	HS_227,HS_228	1	15471992	25039867	1p36.21	1p36.11	rs761301	rs1890450	2970	9567875	137
G2006232_1.1	HS_226,HS_227,HS_228	1	5758772	6100313	1p36.31	1p36.31	rs7521197	rs1295087	<u>98</u>	341541	2
G2006232_6	HS_226,HS_227	9	161424682	161860424	6q26	6q26	rs606817	rs12111173	142	435742	4
G2006232_8	HS_226,HS_227	∞	11730094	12718500	8p23.1	8p22	rs1692804	rs17832679	110	988406	14
G2006232_8	HS_226,HS_228	∞	11730094	12718500	8p23.1	8p22	rs1692804	rs17832679	110	988406	14
G2006232_8	HS_227,HS_228	∞	5476393	16007395	8p23.2	8p22	rs2733033	rs351554	3952	10531002	91
G2006232_8	HS_226,HS_227	∞	15968147	16645869	8p22	8p22	rs7015682	rs11203801	175	677722	1
G2006232_8	HS_226,HS_228	∞	17261568	17535483	8p22	8p22	rs11203852	rs7010721	165	273915	4
G2006232_8	HS_226,HS_227	∞	19919169	20484131	8p21.3	8p21.3	rs10098925	rs1031317	207	564962	ŝ
G2006232_8	HS_226,HS_227,HS_228	∞	11730094	12718500	8p23.1	8p22	rs1692804	rs17832679	110	988406	14
G2006232_8	HS_226,HS_227,HS_228	∞	15968147	16007395	8p22	8p22	rs7015682	rs351554	6	39248	1
G2006232_11.1	HS_226,HS_228	11	1075747	1272858	11p15.5	11p15.5	rs2856111	rs4963059	35	197111	2
G2006232_17	HS_226,HS_227	17	953737	1177002	17p13.3	17p13.3	rs2257222	rs1472180	67	223265	2
G2006232_17	HS_226,HS_227	17	5951088	6211023	17p13.2	17p13.2	rs1972498	rs7217252	109	259935	1
G2006232_17	HS_226,HS_227	17	6456609	6726112	17p13.2	17p13.1	rs11078637	rs9899579	114	269503	6
G2006232_17	HS_226,HS_227	17	10899108	11229733	17p12	17p12	rs2321321	rs205027	119	330625	1
G2006232_22	HS_226,HS_227	22	48936043	49030818	22q13.32	22q13.32	rs4823802	rs761783	34	94775	1
LOH, Loss Of Hete	erozygosity; LOH_ID, Identité	de la régi	ion de LOH ; T	samples, Ech	antillon tumo	ral; Chrom, C	hromosome; S	tartpos/Starto	:yto/Star	tSNP, Début de	ela
position génomic	que/cytologique/SNP des ha	olotypes c	conservés; En	idpos/Endcyto	o/EndSNP, Fin	de la positio	n génomique/(:ytologique/SI	NP des h	aplotypes con	servés;
Nsnp, Nombre de	SNPs contenus dans chaqu	e haplotyp	oe; Length.bp,	taille de l'ha	plotype en pa	ire de bases;	Ngenes, Nomb	re de gènes co	ontenus (dans chaque h	aplotype.
Tableau 9. Ca	aractéristiques c	les 22 ha	aplotypes c	onservés d	étectés poi	ur les deux	cancers du	sein analys	és dans	s la famille G	2009159
------------------	----------------------	-------------	----------------	----------------	------------------	----------------	------------------	-----------------	-----------	-----------------	------------
	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G2009159_4	HS_237,HS_239	4	6697766	9930962	4p16.1	4p16.1	rs3822262	rs881971	930	3233196	28
G2009159_4	HS_237,HS_239	4	36721478	55684426	4p14	4q12	rs6843202	rs2646356	3297	18962948	82
G2009159_4	HS_237,HS_239	4	83431048	91829443	4q21.22	4q22.1	rs10016201	rs7676836	1836	8398395	51
G2009159_4	HS_237,HS_239	4	92076288	94702001	4q22.1	4q22.2	rs10016028	rs10084947	375	2625713	2
G2009159_4	HS_237,HS_239	4	106263681	107807776	4q24	4q25	rs6823995	rs11940657	291	1544095	8
G2009159_4	HS_237,HS_239	4	108396913	109474255	4q25	4q25	rs17501731	rs1017432	196	1077342	7
G2009159_4	HS_237,HS_239	4	109553331	110491665	4q25	4q25	rs2880787	rs6831048	160	938334	S
G2009159_4	HS_237,HS_239	4	111811620	117795603	4q25	4q26	rs537772	rs10033291	1111	5983983	20
G2009159_4	HS_237,HS_239	4	117803013	122308022	4q26	4q27	rs11098361	rs692496	931	4505009	19
G2009159_4	HS_237,HS_239	4	122624065	132118585	4q27	4q28.3	rs7680787	rs17447684	1867	9494520	31
G2009159_4	HS_237,HS_239	4	132409579	141012918	4q28.3	4q31.1	rs351096	rs7691693	1656	8603339	13
G2009159_4	HS_237,HS_239	4	178106459	190903688	4q34.3	4q35.2	rs966023	rs2077107	3972	12797229	46
G2009159_6	HS_237,HS_239	9	69384471	72451423	6q12	6q13	rs1033520	rs1328693	721	3066952	12
G2009159_6	HS_237,HS_239	9	167131911	167544709	6q27	6q27	rs12660310	rs3798315	149	412798	4
G2009159_9	HS_237,HS_239	6	133057987	133284310	9q34.11	9q34.11	rs10988681	rs10793975	29	226323	0
G2009159_9	HS_237,HS_239	6	133350608	139257411	9q34.11	9q34.3	rs1215972	rs10870202	2021	5906803	87
G2009159_10	HS_237,HS_239	10	83301681	83978695	10q23.1	10q23.1	rs2224082	rs17099627	213	677014	1
G2009159_15	HS_237,HS_239	15	44984631	45544319	15q21.1	15q21.1	rs12594463	rs2899376	124	559688	6
G2009159_17	HS_237,HS_239	17	35483943	37303198	17q12	17q12	rs17138999	rs12946701	366	1819255	28
G2009159_22	HS_237,HS_239	22	18542311	18893631	22q11.21	22q11.21	rs7288409	rs407895	52	351320	4
G2009159_22	HS_237,HS_239	22	20278471	20736315	22q11.21	22q11.21	rs5748573	rs361764	11	457844	m
G2009159_22	HS_237,HS_239	22	25374503	51211392	22q11.23	22q13.33	rs2103587	rs3888396	8431	25836889	342
LOH, Loss Of Het	terozygosity; LOH_IC	O, Identité	de la région d	e LOH ; Tsamp	oles, Echantil	Ion tumoral;	Chrom, Chromo	osome; Startpo	s/Startcy	yto/StartSNP, D	ébut de la
position génom	ique/cytologique/Sh	vP des hap	lotypes conse	ervés; Endpos	/Endcyto/End	ISNP, Fin de l	a position géno	omique/cytolo	gique/SN	IP des haplotyp	es
conservés; Nsnp	o, Nombre de SNPs c	ontenus d	ans chaque h	aplotype; Leng	gth.bp, taille o	de l'haplotyp	e en paire de ba	ases; Ngenes, N	Vombre o	de gènes conter	us dans
chaque haploty	pe.										

Tableau 10. C	Caractéristiques	des 12 h	aplotypes	conservés (détectés po	ur les deux	cancers du	sein analys	és dan:	s la famille G	2010325
п_нол	TSamples	Chrom	Startpos	Endpos	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
G2010325_9	HS_249,HS_250	6	1793162	1963237	9p24.3	9p24.3	rs7852096	rs10757069	99	170075	0
G2010325_13	HS_249,HS_250	13	22116804	22473593	13q12.11	13q12.11	rs2761909	rs7993238	121	356789	2
G2010325_13	HS_249,HS_250	13	22606401	22788665	13q12.11	13q12.11	rs3117918	rs11148372	115	182264	0
G2010325_13	HS_249,HS_250	13	22875838	23287775	13q12.11	13q12.11	rs9506831	rs314863	125	411937	0
G2010325_13	HS_249,HS_250	13	24495061	24782064	13q12.12	13q12.12	rs7317321	rs4770628	195	287003	2
G2010325_13	HS_249,HS_250	13	27398788	28385757	13q12.13	13q12.2	rs7335408	rs10162043	345	986969	11
G2010325_13	HS_249,HS_250	13	28402332	44574009	13q12.2	13q14.11	rs874497	rs7333269	5069	16171677	83
G2010325_17	HS_249,HS_250	17	13043	2787138	17p13.3	17p13.3	rs17054921	rs7226198	818	2774095	52
G2010325_21	HS_249,HS_250	21	17004225	17661418	21q21.1	21q21.1	rs2823407	rs2823739	161	657193	2
G2010325_21	HS_249,HS_250	21	27763523	28187060	21q21.3	21q21.3	rs2830194	rs2830546	187	423537	1
G2010325_21	HS_249,HS_250	21	28203668	34481036	21q21.3	21q22.11	rs229036	rs11088238	1734	6277368	69
G2010325_21	HS_249,HS_250	21	34875373	48090629	21q22.11	21q22.3	rs7281539	rs9637231	4571	13215256	150
LOH, Loss Of Het	erozygosity; LOH_IC	D, Identité	de la région c	le LOH ; Tsamp	oles, Echantillo	on tumoral; C	hrom, Chromo:	some; Startpos	/Startcyt	to/StartSNP, Dél	out de la
position génomi	ique/cytologique/SN	NP des hap	lotypes const	ervés; Endpos	;/Endcyto/End	SNP, Fin de la	position génoi	mique/cytologi	que/SNP	o des haplotype	s
conservés; Nsnp	o, Nombre de SNPs c	ontenus da	ans chaque h	aplotype; Leng	gth.bp, taille d	e l'haplotype	en paire de ba	ses; Ngenes, No	ombre de	e gènes contenu	s dans
chaque haplotyp	je.										





chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues 'absence de SNP analysé. Les haplotypes conservés sont représentés par une accolade au-dessus des haplotypes tumoraux et correspondent à Visualisation sous Integrative Genomics Viewer (IGV) de deux haplotypes conservés identifiés sur le chromosome 11 [11q14.3 – 11q22.1] et 11q22.1 - 11q22.3] pour les tumeurs du sein HS_151 et HS_154 et d'un haplotype conservé identifié sur le chromosome 16 [16q22.1 -6q22.2] pour les tumeurs du sein HS_153 et HS_154 de la famille G1995037. (a) Représentation du chromosome et de la région du et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises traduisent la partie de l'haplotype qui est identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les haplotypes tumoraux Figure 23: Exemple de visualisation de trois haplotypes conservés pour les trois tumeurs du sein (liées 2 à 2) de la famille G1995037. selon RefSeq Genes.



Figure 24: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G1995042.

L'haplotype conservé est représenté par une accolade au-dessus des haplotypes tumoraux et correspond à la partie de l'haplotype qui est Visualisation sous Integrative Genomics Viewer (IGV) d'un haplotype conservé identifié sur le chromosome 14 [14q21.1 – 14q21.3] pour les tumeurs du sein HS_169 et HS_171 de la famille G1995042. (a) Représentation du chromosome et de la région du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grisées traduisent l'absence de SNP analysé. identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les haplotypes tumoraux selon RefSeq Genes.



Figure 25: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G1995079.

Visualisation sous Integrative Genomics Viewer (IGV) d'un haplotype conservé identifié sur le chromosome 22 [22q12.3] pour les tumeurs du sein HS_176 et HS_178 de la famille G1995079. (a) Représentation du chromosome et de la région du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises traduisent l'absence de SNP analysé. L'haplotype conservé est représenté par une accolade au-dessus des haplotypes tumoraux et correspond à la partie de l'haplotype qui est identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les haplotypes tumoraux selon RefSeq Genes.



Figure 26: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G2002199.

SNP analysé. L'haplotype conservé est représenté par une accolade au-dessus des haplotypes tumoraux et correspond à la partie de es tumeurs du sein HS_219 et HS_221 de la famille G2002199. (a) Représentation du chromosome et de la région du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises traduisent l'absence de l'haplotype qui est identique entre les tumeurs familiales. Ce dernier étant très étendu, seules ses extrêmiés sont ici repésentées. (c) Visualisation sous Integrative Genomics Viewer (IGV) d'un haplotype conservé identifié sur le chromosome 17 [17p13.3 - 17p11.1] pour Représentation des gènes contenus dans les haplotypes tumoraux selon RefSeq Genes.



Figure 27: Exemple de visualisation d'un haplotype conservé pour les trois tumeurs du sein de la famille G2006232.

traduisent l'absence de SNP analysé. L'haplotype conservé est représenté par une accolade au-dessus des haplotypes tumoraux et Visualisation sous Integrative Genomics Viewer (IGV) d'un haplotype conservé identifié sur le chromosome 8 [8p23.1 - 8p22] pour les tumeurs du sein HS_226, HS_227 et HS_228 de la famille G2006232. (a) Représentation du chromosome et de la région du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises correspond à la partie de l'haplotype qui est identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les naplotypes tumoraux selon RefSeq Genes.



du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises 4q28.3 - 4q31.1] pour les tumeurs du sein HS_237 et HS_239 de la famille G2009159. (a) Représentation du chromosome et de la région traduisent l'absence de SNP analysé. Les haplotypes conservés sont représentés par une accolade au-dessus des haplotypes tumoraux et correspondent à la partie de l'haplotype qui est identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les Visualisation sous Integrative Genomics Viewer (IGV) de deux haplotypes conservés identifiés sur le chromosome 4 [4q27 – 4q28.3 et Figure 28: Exemple de visualisation de deux haplotypes conservés pour les deux tumeurs du sein de la famille G2009159. naplotypes tumoraux selon RefSeq Genes.



Figure 29: Exemple de visualisation d'un haplotype conservé pour les deux tumeurs du sein de la famille G2010325.

Visualisation sous Integrative Genomics Viewer (IGV) d'un haplotype conservé identifié sur le chromosome 13 [13q12.2 – 13q14.11] pour les tumeurs du sein HS_249 et HS250 de la famille G2010325. (a) Représentation du chromosome et de la région du chromosome précisément concernée. (b) L'haplotype tumoral est représenté pour chaque échantillon telle une série de lignes verticales bleues et jaunes correspondant aux L'haplotype conservé est représenté par une accolade au-dessus des haplotypes tumoraux et correspond à la partie de l'haplotype qui est SNPs, avec pour génotype respectif "A" et "B" selon la nomenclature Illumina. Les zones grises traduisent l'absence de SNP analysé. identique entre les tumeurs familiales. (c) Représentation des gènes contenus dans les haplotypes tumoraux selon RefSeq Genes. La visualisation de ces résultats sous IGV montre pour ces régions une parfaite concordance des SNPs analysés entre les paires ou les trios de tumeurs familiales. Un exemple est donné pour chaque famille dans les Figures 23 à 29.

Ces régions candidates restent cependant assez étendues, 64Mb [min 15 – max 132] en moyenne par famille et contiennent de nombreux gènes, 484 [min 164 – max 802] en moyenne par famille.

4.1.2 Recherche de variants candidats par Séquençage d'exome

L'approche par génotypage tumoral a permis d'identifier différentes régions candidates dans sept familles. Etant donné la taille de ces régions et le nombre de gènes qu'elles pouvaient contenir, nous avons choisi de les explorer par séquençage d'exome afin d'y rechercher de potentiels variants constitutionnels délétères. Ces exomes ont été réalisés à partir des ADNs leucocytaires des sept cas index de ces familles.

4.1.2.1 Résultats globaux

Pour chaque patient, environ 50 000 000 de séquences ont été générées pour les *reads* 1, et le même nombre pour les *reads* 2 (séquençage *Paired-End*). Le pourcentage d'alignement était de 96% en moyenne, le score de qualité moyen par base de 35 (Qphred), la profondeur moyenne de 75X, la couverture exomique moyenne de 99% à 1X, 97% à 10X et 89% à 25X, l'ensemble témoignant d'un séquençage de bonne qualité.

2700 INDELs et 37 000 SNVs environ ont ainsi été identifiés pour chaque patient.

4.1.2.2 Tri des variants par le logiciel ERIS®

Les variants détectés pour chaque patient étant nombreux, ils ont été triés à l'aide du logiciel ERIS® d'IntegraGen avec les filtres suivants (Figure 30):

 Statut hétérozygote des variants (une transmission autosomique dominante étant suspectée dans nos familles);

INTEGRASEN	Filter results by	variants						
ORIS	Variations 🕜 : 🛛	v IIA						
	Type :	3'UTR	🥑 Mi	ssense		Exonic (indel)		
		5'UTR	No No	nsense		Splice intron		
			- C,	lionymous				
	External Databas	ses:					Quality Control :	
	Keep known SNPs	(dbsnp132) 😢 :	Yes 🥗 Exclude be	enign Polyphen2 😮 :	No 💝		Min depth 😧 : 10	
	EVS variant freq m	ax (%) : 📀	0,5 Integragen	het freq max (%) 😢 :	0,5		Exclude ambiguous ge	enotypes 😗 :
	1000G variant freq	max (%) 😯 :	0,5 Integragen	homo freq max (%) 🔞 :	0,5		Exclude "No Coverage	e" 🔞 : 📃
	Hapmap variant fre	eq max (%) 😢 :	0,5					
	Locus / Gene se	earch :					Other:	
	Position ₀ :	chr14:101356	6010-10728198				Breakpoints only 😮 :	No 👓
	Filler hu conc(c)			Cites by sefere			miRNAs only 😢 :	No 👓
	Pliter by gene(s)			itter by refseq			Exclude Ensembl ₀ :	No 👓
				J		.::		
		Group n°1						
		Status :	HTZ 👓					
		G1995042	2-11:14					
	l							

Figure 30: Procédure de tri des variants dans le logiciel ERIS® d'IntegraGen.

La société IntegraGen a developpé le logiciel ERIS® pour aider à trier les nombreux variants obtenus par patient testé en fonction de leur type (INDEL, non sens, intronique...), de leur qualité (profondeur minimale de lecture), de leur fréquence allélique rapportée dans les bases de données connues (dbSNP, 1000Genomes, HapMap et EVS) et/ou dans la base de données IntegraGen, de leur position génomique et de leur statut hétérozygote (HTZ) ou homozygote (HMZ).

- Variants avec impact délétère clair d'abord (INDELs, SNVs non sens, SNVs au niveau des sites consensus d'épissage) puis les SNVs faux-sens, ainsi que les SNVs synonymes et introniques plus profonds pouvant retentir sur l'épissage;
- Variants avec une fréquence allélique rapportée dans les bases de données 1000Genome, EVS et HapMap <0,5% ;
- Variants avec une fréquence allélique rapportée dans la base d'Integragen au statut hétérozygote (HTZ) et au statut homozygote (HMZ) <0,5% (afin d'éliminer les « variants artefacts » dus à la technique ou à des erreurs d'alignement);
- Variants avec une profondeur de lecture minimale de 10X ;
- Variants présents dans les régions candidates précédemment identifiées par génotypage tumoral pour chaque famille.

En appliquant ces filtres, un total de 91 variants dans 80 gènes différents a été généré pour les sept cas index. On dénombre parmi eux, 43 variants faux sens, 28 variants introniques, 18 variants synonymes et deux duplications en phase.

En revanche, aucun variant clairement délétère, interrompant le cadre de lecture ou retentissant sur l'épissage, n'a été observé dans les régions candidates. Les deux seuls variants « INDELs » rapportés étaient des duplications en phase, respectivement de 1 et 5 triplets de nucléotides, ne les faisant pas considérer comme des mutations ayant un retentissement majeur sur la protéine. De plus, la première n'était pas si rare car retrouvée sur plusieurs échantillons (5/20) de patients non apparentés appartenant à deux projets de recherche différents mais analysés dans le même temps et rapportée avec une fréquence de 25% et 28% respectivement dans les bases ESP (pour *Exome Sequencing Project*) et ExAC (pour *Exome Aggregation Consortium*). La deuxième duplication était en revanche plus rare car rapportée uniquement dans la base ExAC et sans fréquence associée. Toutes les deux cependant, concernaient des gènes, respectivement *C21orf58* et *GPR78*, dont la fonction protéique ne semblait pas liée à la tumorigénèse. *C21orf58* est effectivement un « open reading frame gene » ou gène du cadre de lecture ouvert du chromosome 21 et le gène *GPR78* serait plutôt impliqué dans les troubles bipolaires et la schizophrénie [391]. L'ensemble de ces résultats est résumé par patient et par région candidate dans les Tableaux 11 à 17.

III:12	
patiente G1995037_	
ntifiés chez la	
es conservés idei	
dans les haplotyp	
uençage d'exome (
tés par ség	í
ableau 11. Variants detec	

Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Type	Align GVGD	PolyPhen-2	SIFT	Mutation	Splice
									Taster	
chr11: 91320978 - 100633303	AMOTL1	chr11:94602385	2511C>T	= (p.His837His)	Synonyme				2	Vo effect
chr11: 101594826 - 107841291 -										
chr11: 120379057 - 131077439 h	NFRKB	chr11:129739959	3036A>C	= (p.Thr1012Thr)	Synonyme				2	Vo effect
chr16: 70779144 - 71350049										
chr16: 75840629 - 76905308										
chr16: 81989484 - 82451159 -										
11	00 T T T T O					Corrigin Distriction of the second				4

(-) variant non detecté; C0, C15, C25, C45 and C65, Interference avec la fonction protéique du moins probable (C0) au plus probable (C65); Pr D, Probably damaging; Po D, Possibly damaging; B, Benign; D, Deleterious; T, Tolerated; DC, Disease causing, Pol, Polymorphism; NA, Non annoté; p, p-value; No effect, pas de retentisement prédit du variant sur l'épissage.

14	
≓ı	
5	
95(
19	
e	
ent	
ati	
ар	
N	
ç	
és	
Itif	
h	
i s	
Ž	
nse	
8	
Sec	
tyl	
blq	
ha	
es	
Sug	
ğ	
Ĕ	
š	
P	
age	
Śu	
au	
sé,	
par	
és	
ect	
let	
ţ	
ian	
Var	
2	
5	
ea	
ab	
	l

Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Type	Align GVGD	PolyPhen-2	SIFT	Mutation	Splice
									Taster	
chr14: 39821153 - 47633066	•									
chr14: 54186083 - 55583424	•									
080102010 101356010 107381080	AHNAK2	chr14:105418694	3094C>G	Leu1032Val	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B (0.00)	T (1)	Pol (p:1)	
006107/01 - 010066101 - 10/ 201000	TMEM121	chr14:105995405	234G>C	= (p.Leu78Leu)	Synonyme				~	Vo effect
(-) variant non detecté; C0, C15, C2	5, C45 and C65	i, Interference avec la	fonction protéique d	du moins probable (CO)	au plus proba	ble (C65); Pr D, Probably damag	ging; Po D, Possib	ly dama	ting; B, Benigr); D,
Deleterious; T, Tolerated; DC, Dise	ease causing,	, Pol, Polymorphism; N	A, Non annoté; p, p-	/alue; No effect, pas de	e retentisemer	it prédit du variant sur l'épissa	ge.			

Haplotype cons	ervé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Tvne	Align GVGD	PolvPhen-2	SIFT	Autation	Solice
										Taster	
chr6: 61891118 - 661	164344										
		UFL1	chr6:96997599	1742T>A	lle581Asn	Faux sens	C45 (GV: 28.68 - GD: 133.10)	Po D (0.948) 1	D (0) C	0C (p:1)	
		BCLAF1	chr6:136599748	271G>A	Gly91Ser	Faux sens	C0 (GV: 109.55 - GD: 46.81)	Pr D (1)	r (0.2)	0C (p:1)	
		MAP3K5	chr6:136926510	2522-8_2522-6dup	ۍ	Intronique				2	Vo effect
721 CT10CT00-3442	0710100	GJE1	chr6:142455821	380C>T	Ala127Val	Faux sens	C0 (GV: 232.67 - GD: 35.66)	B (0.006) I	0 (0.04) N	AN M	
VT - 7/160/06 000	0/hcten	LRP11	chr6:150147414	1334C>T	Pro445Leu	Faux sens	C65 (GV: 0.00 - GD: 97.78)	Pr D (1) 1	0 (0) C	0C (p:1)	
		TCP1	chr6:160201543	1029A>G	= (Ala343Ala)	Synonyme				2	Vo effect
		IGF2R	chr6:160468175	2052-16C>T	ۍ	Intronique				2	Vo effect
		PDCD2	chr6:170886648	1034A>G	= (*345*)	Synonyme				~	No effect
		BIRC3	chr11:102196031	791A>T	Asn264Ile	Faux sens	C15 (GV: 74.82 - GD: 89.28)	Po D (0.872) 1	0 (10.0) C	IA	
		DYNC2H1	chr11:103026053	3574-7del	ۍ.	Intronique				2	Vo effect
		DYNC2H1	chr11:103040982	5114T>C	Leu1705Pro	Faux sens	C65 (GV: 4.86 - GD: 95.38)	Pr D (0.998)	0 (0)	0C (p:1)	
chr11: 90070103 - 1	31888483	ATM	chr1:108202225	7570G>C	Ala2524Pro	Faux sens	C25 (GV: 0.00 - GD: 26.87)	Pr D (1) 1	0 (0) C	0C (p:1)	
		DLAT	chr11:111896317	121G>A	Ala41Thr	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B (0.011) [D (0.03) F	ol (p:1)	
		SIK3	chr11:116729374	2663C>T	Ala888Val	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B(0.007)	r (0.39) F	ol (p:1)	
		AMICA1	chr11:118083270	50C>T	Ser17Phe	Faux sens	C25 (GV: 73.35 - GD: 113.73)	B (0.006) I	D (0.02) F	ol (p:1)	
chr22: 22710758 - 23	3258369										
chr22: 34799261 - 39	5964722	ISX	chr22:35478570	289C>G	Leu97Val	Faux sens	C0 (GV: 234.72 - GD: 16.18)	Pr D (1) [D (0) C	C (p:1)	
chr22: 39575848 - 41	1415644	SGSM3	chr22:40803845	1577G>A	Arg526GIn	Faux sens	C0 (GV: 241.31 - GD: 0.00)	Po D (0.914) 1	D (0) C	0C (p:1)	
Ch. OBCAOTAN CONAS	240636	CELSR1	chr22:46761267	8415C>T	= (Ser2805Ser)	Synonyme				~	Vo effect
4 - 00000000000000000000000000000000000	OCCOTC	CELSR1	chr22:46795716	5310C>T	= (p.Thr1770Thr)	Synonyme				~	Vo effect
(-) variant non detecté	é; CO, C15, C	25, C45 and Ct	55, Interference avec l	la fonction protéique o	tu moins probable (CO)	au plus prob	able (C65); Pr D, Probably damag	ing; Po D, Possib	iy damagi	ng; B, Benigr	η; D,

Tableau 13. Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G1995079_III:1

Deleterious; T, Tolerated; DC, Disease causing, Pol, Polymorphism; NA, Non annoté; p, p-value; No effect, pas de retentisement prédit du variant sur l'épissage.

Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Type	Align GVGD	PolyPhen-2	SIFT	Mutation	Splice
									Taster	
chr11: 71626089 - 76301375	CLPB	chr11:72004653	1882C>T	Arg628Cys	Faux sens	C0 (GV: 241.96 - GD: 80.92)	Pr D (1)	D (0,01) N	A	Vo effect
	CCDC90B	chr11:82985658	426+24C>T	~ .	Intronique				~	Vo effect
	ME3	chr11:86175129	919+1003_919+1004ins18	ۍ.	Intronique				2	Vo effect
	ME3	chr11:86175167	919+964_919+966del	ç.	Intronique				2	No effect
chr11: 77855357 - 134381569	T MMP1	chr11:102661228	1325C>T	Thr442Ile	Faux sens	C0 (GV: 243.26 - GD: 55.02)	B (0.255)	D (0.02) PG	(p:0.989) lo	
	TECTA	chr11:121000844	2865C>T	= (p.Asp955Asp)	Synonyme				2	Vo effect
	DDX25	chr11:125786983	875G>A	Arg292GIn	Faux sens	C0 (GV: 126.59 - GD: 0.00)	Pr D (1)	T (0.29) D	c (p:1)	
	NFRKB	chr11:129753941	816+24C>T	?	Intronique				~	No effect
chr13: 104352252 - 10494113	3-									
	ABR	chr17:.934983	1791+18306_1791+183076	15	Intronique				~	No effect
	ABR	chr17:994886	531+19G>T	ۍ.	Intronique				2	Vo effect
chr17-13043 - 22213008	MYO1C	chr17:1371342	2836G>A	Ala946Thr	Faux sens	C0 (GV: 155.47 - GD: 0.00)	Pr D (0.964)	T (0.07) PG	ol (p:1)	
	MYO1C	chr17:1378011	1717-27A>C	ۍ.	Intronique				2	Vo effect
	PITPNM3	chr17:6380447	987G>A	= (p.Leu329Leu)	Synonyme				2	Vo effect
	ASGR1	chr17:7077097	757G>A	Glu253Lys	Faux sens	C55 (GV: 0.00 - GD: 56.87)	Pr D (1)	D (0) D	C (p:1)	
(-) variant non detecté; C0, C15, (225, C45 and C	265, Interference avec l	la fonction protéique du moi	uld ne (00) au plu:	s probable (C	65); Pr D, Probably damaging; Po	o D, Possibly da	imaging; B, E	Benign; D, Del	eterious;
T, Tolerated; DC, Disease causir	ng, Pol, Polym	orphism; NA, Non ann	oté; p, p-value; No effect, pas	de retentisement préd	it du variant :	sur l'épissage.				

Tableau 14. Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2002199_11:18

121

Haplotype conserve	Gene	Position (g.)	Nucleotide (c.)	Acide Anmine (p.)	Type	Align GVGD	PolyPhen-2	SIFT Mut	tation	Splice
								Ta	ster	
chr1: 5709833 - 6141325										
	ATP13A2	chr1:17328838	588C>T	= (p.Asp196Asp)	Synonyme				ž	o effect
	UBR4	chr1:19422030	13626+9C>G	ç.,	Intronique				ž	o effect
	UBR4	chr1:19436890	11961G>A	= (p.Glu3987Glu)	Synonyme				ž	o effect
	PLA2G2A	chr1:20305235	32T>C	Met11Thr	Faux sens	C0 (GV: 174.65 - GD: 0.00)	B (0.107)	T (0.14) Pol (_k	0:1)	
chr1.1E471007_3E02067	VWA5B1	chr1:20644149	690T>A	= (p.Arg230Arg)	Synonyme				ž	o effect
10066067 - 7661 /+61 -T IID	VWA5B1	chr1:20644157	698G>T	Cys233Phe	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Po D (0.886)	D (0) DC (p	(676)	
	RAP1GAP	chr1:21977906	44+366_44+353del	ç.,	Intronique				ž	o effect
	EPHA8	chr1:22915816	1315+117C>T	ç.,	Intronique				ž	o effect
	KDM1A	chr1:23346213	124G>A	Ala42Thr	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B (0.056)	T (0.89) Pol (J	0:0.917)	
	CNR2	chr1:24201125	983C>T	Ala328Val	Faux sens	C0 (GV: 243.92 - GD: 0.00)	B (0.026)	T (0.32) Pol (_F	0:1)	
chr6: 161424682 - 161860424										
rhr8. 5/76202 _ 166/5860	RP1L1	chr8:10467627	3981A>G	= (p.Thr1327Thr)	Synonyme				ž	o effect
	BLK	chr8:11410922	473-1330G>A	5	Intronique				N	o effect
chr8: 17261568 - 17535483										
chr8: 19919169 - 20484131										
chr11: 1075747 - 1272858										
chr17: 953737 - 1177002										
chr17: 5951088 - 6211023										
chr17: 6456609 - 6726112										
chr17: 10899108 - 11229733										
chr22: 48936043 - 49030818										
(-) variant non detecté; C0, C15,	c25, C45 and C	65, Interference avec	la fonction protéique o	lu moins probable (CO)	au plus prob	able (C65); Pr D, Probably dama	ging; Po D, Possi	bly damaging; B,	, Benign; D,	

Tableau 15. Variants detectés par séquençage d'exome dans les haplotypes conservés identifiés chez la patiente G2006232 II:26

Deleterious; T, Tolerated; DC, Disease causing, Pol, Polymorphism; NA, Non annoté; p, p-value; No effect, pas de retentisement prédit du variant sur l'épissage.

Tableau 16. Variants dete	ictés par sé	iquençage d'exon	ne dans les haplot	ypes conservés ide	entifiés chez	la patiente G2009159_II:2:			
Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Type	Align GVGD	PolyPhen-2 SIF	r Mutation	Splice
								Taster	
C300600 3375033 .bada	MRFAP1	chr4:6711290	67G>A	Asp23Asn	Faux sens	C0 (GV: 108.93 - GD: 0.00)	Pr D (0.989) T (0.4	 Pol (p:0.994) 	
7050665 - 0077600 - 1110	GPR78	chr4:8582999	290_304dup15	Val97_Ser101dup	Dup In-frame				
	NWD2	chr4:37446432	2822A>G	Lys941Arg	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Po D (0.816) T (0.0	7) DC (p:1)	
	TLR1	chr4:38798363	2090T>C	Leu697Ser	Faux sens	C65 (GV: 0.00 - GD: 144.08)	Pr D (1) D (0)	DC (p:1)	
	APBB2	chr4:40829165	1716C>T	= (p.Leu572Leu)	Synonyme				No effect
	SLC30A9	chr4:42003677	154C>T	Pro52Ser	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Pr D (1) T (0.9	6) Pol (p:1)	
chr4: 36721478 - 55684426	CORIN	chr4:47663736	1727A>G	Tyr576Cys	Faux sens	C0 (GV: 191.93 - GD: 94.76)	B (0.433) D (0.0	(5) Pol (p:0.915)	
	CORIN	chr4:47695087	813G>C	Arg271Ser	Faux sens	C0 (GV: 205.85 - GD: 0.00)	B(0) T (0.4	 8) Pol (p:1) 	
	FRYL	chr4:48533314	6762G>A	= (p.Ala2254Ala)	Synonyme				No effect
	FRYL	chr4:48548107	5256C>G	= (p.Thr1752Thr)	Synonyme				No effect
	SPATA18	chr4:52936094	518+12del	د.	Intronique				No effect
chr4: 83431048 - 91829443	HERC3	chr4:89625388	2797A>T	Met933Leu	Faux sens	C0 (GV: 231.89 - GD: 14.30)	Pr D (0.997) T (0.1	(6) DC (p:1)	
chr4: 92076288 - 94702001									
chr4: 106263681 - 107807776									
chr4: 108396913 - 109474255									
chr4: 109553331 - 110491665									
	NEUROG2								
chr4: 111811620 - 117795603		chr4:.113436014	618C>G	Asp206Glu	Faux sens	C0 (GV: 89.39 - GD: 0.00)	Pr D (0.996) T (0.7) Pol (p:0.604)	
	ANK2	chr4:114095596	109C>T	Arg37Cys	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Pr D (1) D (0)	DC (p:1)	
chr4: 117803013 - 122308022									
	METTL14	chr4:119610564	195G>A	= (p.Lys65Lys)	Synonyme				No effect
chr4: 122624065 - 132118585									
chr4: 132409579 - 141012918									
1781064E0 100002689	C4orf47	chr4:186353337	302G>A	Ser101Asn	Faux sens	C45 (GV: 0.00 - GD: 46.24)	Po D (0.566) D (0)	Pol (p:1)	
000006061 - 6650010/T - 110	FAT1	chr4:187540095	7645T>C	Phe2549Leu	Faux sens	C15 (GV: 0.00 - GD: 21.82)	B (0.107) D (0)	DC (p:1)	
chr6: 69384471 - 72451423									
chr6: 167131911 - 167544709									
chr9: 133057987 - 133284310									
11773C0C1 - 90303CCC1 .0442	SETX	chr9:135204506	2479A>G	Lys827Glu	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Po D (0.734) T (0.5	6) Pol (p:1)	
TT+/0760T - 000000000T -6110	RXRA	chr9:137314214	910+563C>T	5	Intronique				No effect
(-) variant non detecté; C0, C15,	C25, C45 and C	65, Interference avec	la fonction protéique o	Ju moins probable (CO)	au plus proba	ole (C65); Pr D, Probably damagin	g; Po D, Possibly dama	ging; B, Benign; D,	

Tableau 16. Variants det	ectés par sé	quençage d'exon	ie dans les haploty	ypes conservés ide	intifiés chez	la patiente G2009159_II:2	5 (Suite)	
Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.)	Type	Align GVGD	PolyPhen-2 SIFT Mutati	on Splice
							Taste	
chr10: 83301681 - 83978695	,							
chr15: 44984631 - 45544319								
chr17: 35483943 - 37303198								
chr22: 18542311 - 18893631	USP18	chr22:18655932	907G>A	Glu303Lys	Faux sens	C0 (GV: 126.06 - GD: 0.00)	Po D (0.851) T (0,25) Pol (p:0.	871)
chr22: 20278471 - 20736315								
	CCDC117	chr22:29182235	761C>T	Ser254Leu	Faux sens	C0 (GV: 184.71 - GD: 0.00)	B (0) T (0,9) Pol (p:1)	
	RFPL1	chr22:29837600	443G>A	Arg148GIn	Faux sens	C0 (GV: 241.31 - GD: 0.00)	Po D (0.926) T (0,53) Pol (p:1)	
	NF2	chr22:30077552	1699G>A	Asp567Asn	Faux sens	C0 (GV: 130.98 - GD: 0.00)	B (0.110) T (0,37) DC (p:0.5	(86)
	SH3BP1	chr22:38037118	60-15del	ç.,	Intronique			No effect
chr22: 25374503 - 51211392	PLAZGD	CNr22:385966668	209431>C	. .	Intronique			No ettect
	SHISA8	chr22:42306100	863C>G	Pro288Arg	Faux sens	C0 (GV: 353.86 - GD: 0.00)	T (0.32) NA	No effect
	CRELD2	chr22:50316989	919+24_919+37dup	٥.	Intronique			No effect
	PPP6R2	chr22:50857914	845+23G>A	ç.,	Intronique			No effect
	SHANK3	chr22:51169556	5060T>A	Phe1687Tyr	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B (0) T (1) Pol (p:0.)	(866
(-) variant non detecté; C0, C15, Deleterious: T Tolerated: DC D	C25, C45 and C	65, Interference avec	la fonction protéique d NA Non annoté- p. p-v	du moins probable (CO) Jalue: No effect pas de	au plus probat	ole (C65); Pr D, Probably damagin : prédit du variant sur l'énissage	g; Po D, Possibly damaging; B, Benig	n; D,

Tableau 17. Variants det	ectés par sé	équençage d'exo	me dans les haplot	types conservés i	dentifiés chez	la patiente G2010325_II:				
Haplotype conservé	Gène	Position (g.)	Nucleotide (c.)	Acide Anminé (p.) Type	Align GVGD	PolyPhen-2	SIFT Mu	itation S	splice
								Ĥ	aster	
chr9: 1793162 - 1963237										
chr13: 22116804 - 22473593										
chr13: 22606401 - 22788665										
chr13: 22875838 - 23287775										
chr13: 24495061 - 24782064										
chr13: 27398788 - 28385757										
chr13: 28402332 - 44574009										
	MY01C	chr17:1384184	628-5C>A	ۍ	Intronique				No	o effect
chr17: 13043 - 2787138	SMG6	chr17:2200527	2151+10C>T	ۍ.	Intronique				No	o effect
	CLUH	chr17:2598498	2490+14G>A	5	Intronique				No	o effect
chr21: 17004225 - 17661418										
chr21: 27763523 - 28187060										
chr21: 28203668 - 34481036	URB1	chr21:33725029	2603G>A	Arg868GIn	Faux sens	C0 (GV: 353.86 - GD: 0.00)	B (0.032)	T (0.62) Pol	(p:1)	
	C21orf58	chr21:47721986	894_896dup	His299dup	Dup In-frame					
	SIM2	chr21:38081762	258+212C>T	٥.	Intronique				No	o effect
	TTC3	chr21:38536460	3278T>G	Val1093Gly	Faux sens	C0 (GV: 353.86 - GD: 0.00)	Po D (0.919)	T (0.06) Pol	(p:1)	
	ERG	chr21:39818724	40-1180C>T	۵.	Intronique				No	o effect
chr21: 34875373 - 48090629	ABCG1	chr21:43711583	1531-25G>A	۵.	Intronique				No	o effect
	PDXK	chr21:45152248	88-1702C>T	۰.	Intronique				No	o effect
	TRPM2	chr21:45837901	3238G>A	Ala1080Thr	Faux sens	C0 (GV: 240.36 - GD: 8.09)	Pr D (1)	T (0.19) DC	(p:1)	
	MCM3AP	chr21:47695076	1998+24G>T	۰.	Intronique				No	o effect
	MCM3AP	chr21:47704577	624C>A	= (p.Thr208Thr)	Synonyme				No	o effect
(-) variant non detecté; 00, C15, (C25, C45 and O	55, Interference avec	la fonction protéique (du moins probable (O	0) au plus probat	ole (C65); Pr D, Probably damagir	ig; Po D, Possibly	damaging; B,	Benign; D,	

4.1.2.3 Les variants candidats

Les variants synonymes et introniques identifiés chez nos sept patients ne présentaient aucun intérêt, les premiers ne modifiant pas la séquence protéique, les seconds étant en dehors des sites consensus d'épissage et les deux n'ayant aucun impact prédit sur l'épissage (Tableaux 11 à 17).

Par contre, parmi les variants faux sens identifiés, 11 ont attiré notre attention car ils étaient prédits délétères par au moins trois sur quatre logiciels de modélisation protéique dont Align GVGD, PolyPhen-2, SIFT et mutation Taster [392-395] (Tableaux 13 à 16). Ces variants concernaient 11 gènes différents : *VWA5B1*, *TLR1*, *ANK2*, *C4orf47*, *UFL1*, *LRP11*, *DYNC2H1*, *ATM*, *ASGR1*, *ISX* et *SGSM3* dont les rôles ou les fonctions biologiques sont brièvement décrits ci-après.

Le gène VWAB1 (von Willebrand factor A domain containing 5B1) code une protéine (5B1) contenant un domaine du facteur de Willebrand de type A. Très peu d'informations sont disponibles le concernant mais des mutations rapportées dans un gène proche, VWA3B, seraient responsables à l'état homozygote, d'ataxie cérébelleuse avec déficit intellectuel [396].

Le gène *TLR1* code une protéine appartenant à la famille des récepteurs *Toll-like*, laquelle joue un rôle fondamental dans la reconnaissance des agents pathogènes et dans l'activation de l'immunité innée [397].

Le gène *ANK2* code une protéine appartenant à la famille des Ankyrines, lesquelles permettent de lier les protéines membranaires intégrales au cytosquelette sous-jacent formé par l'actine et la spectrine. La protéine codée par *ANK2* intervient plus particulièrement dans la stabilité des canaux Na/Ca situés au niveau des cardiomyocytes. Des mutations de *ANK2* sont d'ailleurs responsables d'arythmies cardiaques et du syndrome du QT long de type 4 [398].

Le gène *C4orf47* est un « *Open Reading Frame gene* » du chromosome 4, sans plus d'information à son sujet.

Le gène *UFL1* code une E3 protéine ligase qui permet l'ufmylation (ou conjugaison) de certaines protéines comme DDRGK1. Le rôle de l'Ufmylation n'est pas bien connu mais cette protéine pourrait agir comme un suppresseur de tumeur en bloquant la voie de signalisation NF-kappa-B [399].

Le gène LRP11 code un récepteur de lipoprotéine de basse densité (LDL).

Le gène *DYNC2H1* code une large Dynéine cytoplasmique impliquée dans le transport ciliaire rétrograde et jouant un rôle dans le transport intra-flagellaire. Des mutations dans ce gène sont associées à un large spectre d'anomalies allant de la dyskinésie ciliaire primitive à des anomalies squelettiques plus ou moins complexes [400].

Le gène *ASGR1* code un récepteur transmembranaire asialoglycoprotéique, exprimé exclusivement dans le foie. Ce dernier joue un rôle critique dans l'homéostasie des glycoprotéines dans le sérum en permettant leur endocytose et leur dégradation lysosomale. Une haploinsuffisance *ASGR1* serait par ailleurs associée à un risque réduit de coronaropathie [401].

Le gène *ISX* appartient à la famille des gènes homeobox RAXLX. Il code un facteur de transcription régulant l'expression des gènes dans l'intestin. Ce dernier participerait ainsi au métabolisme de la vitamine A en régulant l'expression de BCO1 dans l'intestin [402].

Le gène *SGSM3 (small G protein signaling modulator 3)* code une protéine modulant la voie de signalisation médiée par les protéines RAP et RAB, ces dernières étant des « small GTPase » [403]. La protéine codée par *SGSM3*, en interagissant avec le suppresseur de tumeur Merlin, jouerait également un rôle dans la suppression de la croissance cellulaire [403;404].

Le gène *ATM*, comme vu précédemment dans le paragraphe 1.4.3.2, code une sérine/thréonine kinase jouant un rôle clé dans le contrôle du cycle cellulaire, la détection et la signalisation des dommages causés à l'ADN et la réparation des cassures double-brin de l'ADN, en

phosphorylant des protéines majeures comme BRCA1, P53 et CHEK2 [277;311]. Ses mutations sont responsables, à l'état homozygote ou hétérozygote composite, du syndrome d'Ataxie-Télangiectasie alors qu'à l'état hétérozygote, elles sont associées à un risque modéré de cancer du sein.

De manière intéressante, la mutation faux sens identifiée ici chez la patiente G1995079_III:1 et nomenclaturée c. 7570G>C (p. Ala2524Pro), a été rapportée précédemment dans deux familles AT finlandaises, en association avec une mutation «*frameshift* » pour la première et à l'état homozygote pour la deuxième [405;406]. Dans ces deux familles, plusieurs cas de cancers du sein ont été observés chez les porteurs hétérozygotes de la mutation c.7570G>C. Cette même mutation a de plus été identifiée à l'état hétérozygote dans au moins trois familles finlandaises, non AT, avec agrégation familiale de cancers du sein [406]. L'ensemble de ces deux observations était donc en faveur du caractère délétère de la mutation faux sens et de son implication dans le risque de cancer du sein.

Ainsi, la mutation faux sens touchant le gène *ATM* constituait le seul véritable variant candidat de prédisposition héréditaire au cancer du sein, la pathogénicité prédite pour les autres variants détectés étant effectivement difficile à démontrer. De plus, la majorité de ces derniers concernait des gènes dont l'implication dans la cancérogénèse mammaire n'était pas évidente.

4.2 Etude 2 : A propos de 11 Formes familiales de cancer du sein associées à une mutation constitutionnelle du gène *ATM*

La mise en évidence d'une mutation délétère du gène *ATM* dans une famille avec récurrence familiale de cancer du sein, très suspecte d'une hérédité sous-jacente, constituait un résultat intéressant dans la mesure où le rôle d'*ATM* dans la prédisposition héréditaire au cancer du sein, a longtemps été un sujet discuté et controversé [299;303-308;407]. En effet, des études épidémiologiques menées dans des familles AT ont montré dans les années 1990 que les hétérozygotes présentaient un risque de cancer du sein augmenté [299;300;302]. Ce risque était cependant modéré et donc insuffisant pour expliquer des formes familiales de cancer du

sein, avec un risque relatif compris entre 2 et 4 selon les études précédentes, contre 11 et 12 pour les mutations *BRCA1* et *BRCA2* [408]. Ce n'est qu'en 2006, qu'une plus large étude portant cette fois sur des cas de cancer du sein familiaux a révélé que le risque de cancer du sein était augmenté chez les hétérozygotes *ATM* non AT, mais que ce risque restait modéré avec un RR à 2,37 (p = 0,0003) [174]. Depuis, peu d'études ont étudié et précisé le rôle d'*ATM* dans ce contexte.

Parallèlement à l'identification de cette mutation, nous nous sommes aperçu au laboratoire d'Oncogénétique de l'Institut Bergonié, que l'identification de mutations incidentes pour le gène *ATM* dans le cadre du diagnostic moléculaire réalisé par NGS ciblé, était clairement plus fréquente lorsque l'indication d'analyse était le diagnostic de formes familiales de cancer du sein plutôt que le diagnostic de formes familiales de cancer digestif ou autres syndromes de prédisposition comme la maladie de Cowden ou le syndrome de Gorlin (données non publiées).

L'ensemble de ces observations nous a conduits à poursuivre nos investigations en sélectionnant dix familles supplémentaires avec agrégation familiale de cancers du sein, associées à une mutation constitutionnelle délétère du gène *ATM* identifiée au laboratoire chez le cas index. L'objectif principal de cette nouvelle étude était d'explorer les pertes d'hétérozygotie au locus d'*ATM* dans les tumeurs du sein familiales afin d'évaluer, selon le modèle de Knudson, son implication dans la tumorigénèse de ces tumeurs. Le deuxième objectif était également d'étudier la ségrégation des mutations *ATM* identifiées avec les atteintes familiales lorsqu'elles étaient disponibles. Cette étude a fait l'objet d'un article, présenté ci-après.

Genes, Chromosomes & Cancer

Frequent wild-type allele loss argues in favor of the involvement of ATM germline mutations in a subset of breast cancer families

Journal:	Genes, Chromosomes and Cancer
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Keywords:	Familial breast cancer, ATM, SNP array, NGS, Breast cancer risk



John Wiley & Sons

Frequent wild-type allele loss argues in favor of the involvement of *ATM* germline mutations in a subset of breast cancer families

Virginie Bubien^{1,2}, Françoise Bonnet^{1,2}, Jennifer Dupiot-Chiron², Emmanuelle Barouk-Simonet², Natalie Jones^{1,2}, Aurélien de Reynies³, Gaëtan MacGrogan^{1,4}, Nicolas Sevenet^{1,2}, Eric Letouzé³, Michel Longy^{1,2}

1/INSERM U1218, Université de Bordeaux, Bordeaux France

2/Cancer Genetics Unit, Institut Bergonie, Bordeaux France

3/Programme CIT, Ligue contre le cancer, Paris, France

4/BioPathology Department, Institut Bergonié, Bordeaux, France

Correspondence to:

Michel Longy

Cancer genetics unit,

Institut Bergonié,

229, cours de l'Argonne,

33000 Bordeaux

m.longy@bordeaux.unicancer.fr

tel: 05 56 33 04 39 - fax: 05 47 30 60 65

Keywords: Familial breast cancer, ATM, SNP array, NGS, Breast cancer risk

Abstract:

Familial breast cancers (BCs) account for 10-20% of all diagnosed BCs, yet only 50% of such tumors arise in the context of a germline mutation in known tumor suppressor genes such as BRCA1 or BRCA2. The vast genetic heterogeneity which characterizes BRCAx families makes grouped studies impossible to perform. Next generation sequencing (NGS) techniques, however, allow individual families to be studied in order to identify private mutations. Single nucleotide polymorphism (SNP) arrays allow the detection of conserved haplotypes within recurrent regions of loss of heterozygosity, common to several familial tumors, therefore identifying genomic loci likely to harbor a germline mutation in cancer predisposition genes. The combination of both exome sequencing and SNP arrays for a series of familial BC, revealed a germline ATM mutation associated with a loss of the wild-type allele in a BRCAx family. The analysis of additional breast tumors from ten BC families in which a germline ATM mutation had been identified revealed a high frequency of wild-type allele loss in these tumors (>80% compared to the 20% expected in sporadic BC; p <0.001). This result argues strongly in favor of the involvement of ATM in the carcinogenesis of these tumors as a tumor suppressor gene and suggests that germline ATM mutations are involved in a subset of familial BC.

Introduction

Breast cancer (BC) is the most common cancer among women, with one in nine women diagnosed in their lifetime in the developed world. Although the majority of BC cases are sporadic, 10-20% of BC patients have a familial history with multiple affected family members (Rahman & Stratton, 1998). The early detection of BC reduces the morbidity and mortality of the disease therefore the diagnosis of hereditary BC predisposition may help improve the disease management. In the 1990s, BRCA1 and BRCA2germline mutations in genes were identified using linkage analysis and are associated with a high BC risk (Miki et al., 1994; Wooster et al., 1995). Using the same strategy or through a candidate gene involved in DNA repair approach, mutations in other high and moderate penetrance cancersusceptibility genes have been identified (Walsh & King, 2007). However, a significant proportion (>50%) of the families (known as BRCAx), are not associated with mutations in BRCA1, BRCA2 or other known BC susceptibility genes (Stratton & Rahman, 2008). The difficulty in identifying new BC susceptibility genes led to the idea of a vast genetic heterogeneity, with private mutations in BRCAx families. Thus, the identification of susceptibility genes is now based on the study of small individual families which does not allow classical gene localization strategies such as linkage analysis to be used.

Recently, Letouzé et al. described a Single nucleotide polymorphism (SNP) arraybased approach allow the detection of conserved haplotypes within recurrent regions of loss of heterozygosity (LOH), common to several tumors, therefore identifying genomic loci likely to harbor a founder germline mutation in cancer predisposition genes (Letouze et al., 2012). Tumors from patients with an inherited germline mutation in a tumor suppressor gene frequently show loss of the wild-type (wt) allele, revealing the mutant phenotype, in accordance with Knudson's two-hit model (Knudson, 2001). Germline mutations are therefore often located in regions of LOH and this has been used to map cancer predisposition genes. LOH results in the retention of only one of the two chromosomes and tumor genotypes obtained by SNP array provide the haplotype of the remaining chromosome. In the present study, we used this approach for two BC from sisters of a BRCAx family in order to detect conserved haplotypes common to both tumors and to identify a pathogenic germline mutation in these candidate regions using exome sequencing. The characterization of an *ATM* mutation prompted us to explore the status of both *ATM* alleles in an additional series of ten familial BC from patients with a germline *ATM* mutation.

Patients and methods

Patients and samples

Breast tumor samples from two individuals of a *BRCAx* family (G1995079) were analyzed by SNP array in order to generate tumor haplotypes. A blood sample from one of these patients (patient III:1) underwent exome sequencing. The pedigree is represented in Figure 1 and the familial BC characteristics are reported in Table 1. An additional 17 breast tumor samples from ten BC families in which index case *ATM* germline mutations had been identified, were screened for mutations in a gene panel using NGS. The pedigrees of these patients are provided in the online supplementary figures (S1-S10).

All index cases of the selected families had undergone molecular genetic diagnostic testing in the context of a suspected breast/ovarian cancer predisposition. No *BRCA1* nor *BRCA2* germline mutations had been identified, with the exception of two patients (G2014701_III:10 and G2009467_III:11) who harbored a *BRCA1* and *BRCA2* mutation, respectively. All patients agreed to the use of their samples for research purposes, in compliance with the French law on tumor banks (law number 2004-800, French Public Health Code articles L. 1243-4 and R. 1243-61) under authorization number AC-2008-812, which was approved by the Comité de Protection des Personnes and all patients agreed to the use of

their samples for germline genetic analysis in compliance with the French law (law n°2004-800).

DNA from peripheral blood and fresh frozen tumor samples was extracted using a conventional phenol / chloroform method. DNA from formalin-fixed embedded tissues was extracted using the QIAamp® DNA FFPE Tissue Kit (QIAGEN France SAS, Courtaboeuf, France) according to the manufacturer's recommendations.

Tumor SNP array

Breast tumors from patients III:1 and III:2 (family G1995079) were analyzed using Illumina HumanOmniExpress v12 chips, containing 731442 probes. Hybridization was performed by IntegraGen (Evry, France), according to the manufacturer's instructions. Raw fluorescent signals were imported into the Illumina BeadStudio software and normalized, as previously described (Peiffer et al., 2006), to obtain a log R ratio (LRR) and B allele frequency (BAF) for each SNP.

Bioinformatics analysis of the SNP arrays data was performed in collaboration with the CIT program from the French Ligue Nationale Contre le Cancer. Genome-wide copynumber changes and therefore LOH, were identified using the Genome Alteration Print (GAP) method, as previously described (Popova et al., 2009). A genomic segment was considered to have undergone LOH when the minor copy had a copy number of <0.5 and the LOH was classed as significant if it covered >1 Mb. The tumor haplotypes were then inferred from their BAF profiles. In genomic regions of LOH, the BAF profile is either close to 0 (A allele retained) or close to 1 (B allele retained). We used fixed thresholds to reconstruct tumor haplotypes in regions of LOH. SNPs with a BAF of >0.65 were assigned the genotype B, whereas SNPs with a BAF of <0.35 were assigned the genotype A. We considered SNPs with a BAF between these two thresholds in at least one of the tumors to be unreliable, and therefore excluded them from the analysis. Once recurrent regions of LOH had been established and tumor haplotypes within these regions reconstructed, the FounderTracker pipeline using the GERMLINE algorithm was used to detect haplotypes significantly conserved in each candidate region, as previously described (Gusev et al., 2009;Letouze et al., 2012). The data were visualized using the Integrative Genomics Viewer (IGV), Broad Institute, USA (Robinson et al., 2011).

Exome Sequencing

Exome capture using the SureSelect Human All Exon Kit (V5; Agilent, Santa Clara, CA), and paired-end sequencing on an Illumina HiSeq 2000 system (San Diego, CA) were performed by IntegraGen using blood DNA from patient G1995079_III:1. Bioinformatic analysis was based on the Illumina pipeline CASAVA1.8[®] using ELANDv2e[®] alignment algorithms and the hg19 human reference genome. Aligned reads were visualized using IGV® or Alamut® (Interactive Biosoftware, Rouen, France). Annotation of genetic variants was performed using an IntegraGen custom pipeline and the IntegraGen ERIS® software was used for variant sorting. Fifty million sequences were generated for both read 1 and read 2, with 96% alignment and a quality score per base (Qphred) of 35. The overall sequencing coverage of the captured regions was 97% and 88% for a depth of 10X and 25X, respectively, resulting in a mean sequencing depth of 68X per base.

Targeted Next Generation Sequencing and ATM mutation allele ratio evaluation

Tumor DNA samples were pair-end sequenced analyzed for a panel of 25 genes including *ATM*, using the SureSelect Target Enrichment Kit and the MiSeq bench top nextgeneration sequencer (Illumina). Details of the gene panel are shown in supplementary table S1. The BCL files generated by the Illumina MiSeq were converted to FASTQ format by the MSR software (Chang et al., 2012) and aligned to the human reference genome hg19 using the BWA algorithm (Li & Durbin, 2010). Local realignment, PCR duplicate marking, basequality recalibration, and calculation of coverage metrics were performed with GATK (McKenna et al., 2010) and Picard tools (v.1.119, http://broadinstitute.github.io/picard/). Putative somatic insertion-deletions (INDELs) and single nucleotide variants (SNVs) were called from bam files using Somatic Variant Caller[®] developed by Illumina. The variants identified were then annotated by Ensembl VEP (McLaren et al., 2016). Only variants that passed the standard Somatic Variant Caller[®] filters and that concerned the *ATM* gene were considered. Aligned reads and variants were visualized using IGV® or Alamut® (Interactive Biosoftware, Rouen, France).

The average sequencing depth of the targeted region was 630X [min 74X – max 1134X] per base and 99% of the targeted region was covered at a depth of \geq 378X.

For each *ATM* mutation, the allele ratio was calculated as the number of mutated reads divided by the total number of reads at that position. The ratio obtained for the tumor DNA was compared to that of the leucocyte DNA, which had been previously analyzed in a diagnostic setting. A theoretical ratio corresponding to a monosomy of chromosome 11 with loss of the wt allele was calculated according to the histological evaluation of normal cell contamination using the formula: y = (1-x)/2 + x, where y is the expected percentage of mutated alleles and x the percentage of tumor cells.

Sanger Sequencing

Sanger sequencing using standard protocols was performed to confirm the presence of *ATM* mutations detected by NGS in 11 index cases, and to study their segregation with disease in family members. Briefly, DNA was amplified by polymerase chain reaction (PCR) using the AmpliTaq Gold[®] Kit (Life Technologies, Paisley, UK). PCR products were purified

then sequenced using an ABI 3130x1 DNA Analyzer (Life Technologies). Specific primers and PCR conditions are provided in supplementary Table S2.

Results

Shared haplotypes are detected within regions of LOH common to two familial breast cancers

Patients III:1 and III:2 from family G1995079 had developed early onset breast cancer, bilateral disease at 30 and 32 years and unilateral disease at 29 years, respectively. Despite the absence of poor prognosis tumor characteristics (small luminal A invasive ductal carcinoma without axillary node involvement) and an optimal treatment (Table 1), they both underwent metastatic relapses and died from their pathology at 35 and 38 years, respectively. The ages at diagnosis and the family history (Fig. 1) are strongly in favor of a cancer predisposition syndrome. *BRCA1* and *BRCA2* analysis performed for both patients did not detect any mutations. SNP-based CGH array performed for the two tumors allowed the identification of three regions with common LOH on chromosomes 6 [6q11.1 - 6q27], 11 [11q14.3 - 11q25] and 22 [22q11.1 - 22q13.31] (Figure 2). For these regions, tumor haplotypes were reconstructed from the B Allele Frequency profiles and analyzed with FounderTracker to detect conserved chromosome segments. This analysis identified 7 shared haplotypes: two on chromosome 6, one on chromosome 11 and four on chromosome 22. The details of these haplotypes are shown in Figure 3 and Table 2. These regions contain a total of 756 genes.

Candidate germline mutations in the regions of interest include the ATM gene.

In order to identify a pathogenic germline mutation in the candidate regions previously identified by SNP array, we performed Exome sequencing to screen the germline DNA of patient III:1 of family G1995079.

A total of 2766 INDELs and 36504 SNVs were identified. INDELs and SNVs of interest were then selected by the IntegraGen ERIS® software using the following criteria: heterozygous

status; all variant types except 3' and 5'UTR variants; Minor Allele Frequency (MAF) of <0,5% in the 1000Genome, EVS, HapMap and IntegraGen databases; read depth of $\ge10X$ and variants located in the 7 conserved haplotypes previously identified by SNP array.

No INDELs, Stop or Splice (consensus splice sites) variants were identified but 19 SNVs including 4 synonymous, 3 intronic and 12 missense variants were detected (Table 3). Six of the missense variants were predicted to be deleterious or probably damaging in at least three of the four predictive programs Align GVGD, SIFT, PolyPhen-2 and MutationTaster (Kumar et al., 2009; Adzhubei et al., 2010; Schwarz et al., 2010; Tavtigian et al., 2006). These variants were found in six different genes: UFL1, LRP11, DYNC2H1, ATM, ISX and SGSM3 (Table 3). Interestingly, the c.7570G>C; p.Ala2524Pro ATM missense variant had been previously reported in two Finnish ataxia telangiectasia (AT) families, in combination with an ATM frameshift mutation in one family and in a homozygous state in the second family (Allinen et al., 2002). Several heterozygous carriers in the two families had BC (Allinen et al., 2002) and furthermore, this particular heterozygous mutation has been identified in at least three Finnish BC families, arguing in favor of its deleterious effect and its implication in BC risk (Pylkas et al., 2007). Sanger sequencing of tumor DNA from the two BC and germline DNA from both affected sisters (patients III:1 et III:2), an unaffected sister (patient III:3) and their parents (patients II:1 et II:2) confirmed the results of exome sequencing and revealed that the mutated allele was inherited from the father and that the unaffected sister was a non-carrier (Figure 4 and Supp. Figure S11). Sanger sequencing of the tumor DNA showing a reduction of the wt allele compared to the germline DNA, confirming the loss of the wt allele in both tumors (Figure 4).

Loss of wild type allele is a frequent event in BCs associated with deleterious germline ATM deleterious mutations

In order to evaluate the frequency of biallelic inactivation of ATM in BCs occurring in the context of a germline ATM mutation, we selected 10 additional BC families with a germline ATM mutation which had been previously identified in a diagnostic setting. The pedigrees are shown in supplementary figures S1-S10 and details of the breast tumors and germline ATM mutations are provided in Table 4. For one patient with bilateral BC (G2009467_III:11), samples from both tumors were available. In five families, tumor samples were also available from family members other than the index case. In three of these tumors (belonging to two families), Sanger sequencing revealed the presence of the familial ATM mutation (Table 4). In total, 14 tumor DNA samples were analyzed and the germline ATM mutation was found in all cases. The mutated allele was found at a ratio greater than 0.60 in 11 tumors (range: 0.64 - 0.90) and when compared to the allelic ratio observed in germline DNA (mean: 0.48; range: 0.41 -0.54) the data suggest that the wt allele is lost in 11/14 tumors (Table 5). Only 3 tumors (G2010321_II:4; G2009467_III:11(T2); G2011018_III:9) displayed the mutated allele at a proportion similar to that in the germline DNA, independent of the level of normal cell contamination. In the case of a monosomy 11 higher theoretical ratios were expected in these tumors (Table 5).

Overall, for the eleven families in which a germline *ATM* mutation had been detected, loss of the wt allele was observed in 13/16 tumors (81%), which is significantly (Chi_2 test; p = 0.0004) greater than the expected frequency of wt LOH at the *ATM* locus in BCs of 20% (three cases) (Hall, 2005;Rio et al., 1998).

Limited evidence of co-segregation of ATM mutations and breast cancer in hereditary BC families

In six of the eleven families, co-segregation analysis was performed using Sanger sequencing to determine if the germline *ATM* mutation segregated with BC cases. Co-segregation was found in seven patients belonging to three different families but four patients each belonging to a different family did not carry the mutated allele (in accordance with the four expected cases if the *ATM* mutation was unrelated to BC) (Tables 1 and 4).

Discussion

Filtering of variants detected using exome sequencing in order to retain those in loci with conserved haplotypes common to tumors from the same family allowed us to identify an ATM germline mutation in a BRCAx family. Germline ATM mutations cause ataxia-telangiectasia (AT), a rare autosomal recessive syndrome characterized by progressive neurodegeneration, oculocutaneous telangiectasia, immunodeficiency, chromosomal instability, radiosensitivity and a predisposition to childhood cancer, particularly lymphoid malignancies (Chun & Gatti, 2004). Previous study of AT families have revealed that heterozygous mutation carriers are at an increased risk of BC (Thompson et al., 2005;Renwick et al., 2006;Janin et al., 1999;Athma et al., 1996) but this risk is moderate and not sufficient to underlie familial BC, with a relative risk (RR) of only 2-4 compared to a RR of 11-12 for BRCA mutations (Easton et al., 2015). Given these observations, several mutation screening studies have been performed to clarify the role of ATM sequence variation in BC risk, but the initial evidence regarding the role of ATM as a BC susceptibility gene was unclear and contradictory (Athma et al., 1996;Bebb et al., 1999;Bishop & Hopper, 1997;Chen et al., 1998;FitzGerald et al., 1997;Izatt et al., 1999;Olsen et al., 2005;Vorechovsky et al., 1996). A study of familial BCs however, showed an estimated RR of BC of 2.37 (95% confidence interval = 1.51-3.78) in monoallelic carriers (Renwick et al., 2006). More recently, the analysis of hereditary BCs by gene-panel sequencing approaches has identified germline ATM mutations at a frequency in 2-5% in
BRCAx patients (Maxwell et al., 2015;Aloraifi et al., 2015;Byers et al., 2016;Ellingson et al., 2015;Snape et al., 2012) compared with 0.5-1% in the general population (FitzGerald et al., 1997;Renwick et al., 2006), thus indicating an over representation of *ATM* germline mutations in a specific group of familial BC. Similarly, Maxwell et al. performed whole-exome sequencing in families at high risk for breast and/or ovarian cancer. In 389 individuals from 243 families, 2.9% were found to have an *ATM* mutation and interestingly, these mutations were significantly more frequent in the subset of individuals with familial BC or early onset BC than in those with individuals with multiple primary breast and/or ovarian cancer (p < 0.01) (Maxwell et al., 2016). These data are consistent with our results as the 11 families that we report are characterized by an aggregation of familial BC with a mean age at diagnosis of 45 years (range: 29-74 years) for the 19 tumors associated with *ATM* germline mutations, and no ovarian cancer cases were reported in these families. Pancreatic cancer cases however, were present in 3 of the 11 families, and six cases of bilateral BC were documented in 4 of the 11 families, as has previously been reported in heterozygous *ATM* mutation carriers (Roberts et al., 2012;Broeks et al., 2000).

Loss of heterozygosity at the *ATM* locus is not a rare event and has been reported in 40% of sporadic breast tumors (Hall, 2005;Rio et al., 1998). If *ATM* was not the specific target of the deletion, this genomic event should affect both wt and mutated alleles with a similar frequency yet over 80% of tumors in our study had specifically lost the wt allele. This significant difference argues strongly in favor of the involvement of *ATM* as tumor suppressor gene in the carcinogenesis of these tumors.

In two of the families, a *BRCA1* (G2014701) or *BRCA2* (G2009467) germline mutation was associated with the *ATM* mutation. To our knowledge, it has not been established if such double heterozygous carriers have are a higher tumor risk. For tumor T1 from patient G2009467_III:11, a loss of the wt allele was observed both for *ATM* and *BRCA2*.

Conversely, no LOH was observed for either of the two genes in tumor T2 from the same patient (data not shown). For patient G2014701_III:10, the tumor allelic ratios showed not only a loss of the wt *ATM* allele but also loss of the mutated *BRCA1* allele (data not shown), suggesting that the tumorigenesis was not due to the germline *BRCA1* mutation but associated with the *ATM* mutation.

These data indicate that germline inactivating mutations of *ATM* are BC susceptibility alleles involved in a subset of familial BC.

Co-segregation of the mutated *ATM* allele with BC cases was observed in only 3 of the 6 informative families, concordant with previous reports (Renwick et al., 2006;Vorechovsky et al., 1996), which showed little evidence of co-segregation and a modest estimated RR of BC associated with *ATM* mutations (Easton et al., 2015). Interestingly, for family G1995079, the *ATM* mutation which segregated with early onset BC in the two affected sisters was inherited from the father and not from the mother who had presented a BC. The absence of BC in the paternal lineage is due to the small size of this branch and the likely predominance of male over female carriers. The maternal lineage, which was not associated with the *ATM* mutation, shows a notable aggregation of BC with five cases affecting the mother, three aunts and a cousin. These cases were late onset (range: 50 to 71 years), in contrast to the early onset (range: 29 to 32 years) and premature death of the two index cases. This observation could raise the question of a digenic inheritance, explaining the early and severe forms of BC observed in the two sisters which may result from the association of the *ATM* mutated allele with an unidentified BC susceptibility allele that segregates within the maternal lineage.

Acknowledgments

We would like to thank the patients who contributed samples to the study. We thank Nabila Elarouci for SNP array analysis, Bernadette Gastaldello and Delfine Lafon who performed DNA extraction, *ATM* analysis and targeted next generation sequencing, the Bergonie Cancer Institute Biological Resources Center for providing samples, Véronique Brouste and Stéphanie Hoppe who performed statistical analysis and Dr. Ravi Nookala for medical writing services. This work is part of the program "Cartes d'Identité des Tumeurs" (CIT, http://cit.ligue-cancer.net) funded and developed by the "Ligue Nationale contre le Cancer". This work was also funded by the Pyrénées-Atlantiques commitee of the "Ligue Nationale contre le Cancer" and by the Lions Club of Bergerac. VB was funded by the Institut Bergonie.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat.Methods 7:248-249.

Allinen M, Launonen V, Laake K, Jansen L, Huusko P, Kaariainen H, Borresen-Dale AL, Winqvist R. 2002. ATM mutations in Finnish breast cancer patients. J.Med.Genet. 39:192-196.

Aloraifi F, McDevitt T, Martiniano R, McGreevy J, McLaughlin R, Egan CM, Cody N, Meany M, Kenny E, Green AJ, Bradley DG, Geraghty JG, Bracken AP. 2015. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. FEBS J. 282:3424-3437.

Athma P, Rappaport R, Swift M. 1996. Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. Cancer Genet.Cytogenet. 92:130-134.

Bebb DG, Yu Z, Chen J, Telatar M, Gelmon K, Phillips N, Gatti RA, Glickman BW. 1999. Absence of mutations in the ATM gene in forty-seven cases of sporadic breast cancer. Br.J.Cancer 80:1979-1981.

Bishop DT, Hopper J. 1997. AT-tributable risks? Nat.Genet. 15:226.

Broeks A, Urbanus JH, Floore AN, Dahler EC, Klijn JG, Rutgers EJ, Devilee P, Russell NS, van Leeuwen FE, van 't Veer LJ. 2000. ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. Am.J.Hum.Genet. 66:494-500.

Byers H, Wallis Y, van Veen EM, Lalloo F, Reay K, Smith P, Wallace AJ, Bowers N, Newman WG, Evans DG. 2016. Sensitivity of BRCA1/2 testing in high-risk breast/ovarian/male breast cancer families: little contribution of comprehensive RNA/NGS panel testing. Eur.J.Hum.Genet.

Chang CT, Tsai CN, Tang CY, Chen CH, Lian JH, Hu CY, Tsai CL, Chao A, Lai CH, Wang TH, Lee YS. 2012. Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. ScientificWorldJournal. 2012:365104.

Chen J, Birkholtz GG, Lindblom P, Rubio C, Lindblom A. 1998. The role of ataxia-telangiectasia heterozygotes in familial breast cancer. Cancer Res. 58:1376-1379.

Chun HH, Gatti RA. 2004. Ataxia-telangiectasia, an evolving phenotype. DNA Repair (Amst) 3:1187-1196.

Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD. 2015. Gene-panel sequencing and the prediction of breast-cancer risk. N.Engl.J.Med. 372:2243-2257.

Ellingson MS, Hart SN, Kalari KR, Suman V, Schahl KA, Dockter TJ, Felten SJ, Sinnwell JP, Thompson KJ, Tang X, Vedell PT, Barman P, Sicotte H, Eckel-Passow JE, Northfelt DW, Gray RJ, McLaughlin SA, Moreno-Aspitia A, Ingle JN, Moyer AM, Visscher DW, Jones K, Conners A, McDonough M, Wieben ED, Wang L, Weinshilboum R, Boughey JC, Goetz MP. 2015. Exome sequencing reveals frequent deleterious germline variants in cancer susceptibility genes in women with invasive breast cancer undergoing neoadjuvant chemotherapy. Breast Cancer Res.Treat. 153:435-443.

FitzGerald MG, Bean JM, Hegde SR, Unsal H, MacDonald DJ, Harkin DP, Finkelstein DM, Isselbacher KJ, Haber DA. 1997. Heterozygous ATM mutations do not contribute to early onset of breast cancer. Nat.Genet. 15:307-310.

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I. 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Res. 19:318-326.

Hall J. 2005. The Ataxia-telangiectasia mutated gene and breast cancer: gene expression profiles and sequence variants. Cancer Lett. 227:105-114.

Izatt L, Greenman J, Hodgson S, Ellis D, Watts S, Scott G, Jacobs C, Liebmann R, Zvelebil MJ, Mathew C, Solomon E. 1999. Identification of germline missense mutations and rare allelic variants in the ATM gene in early-onset breast cancer. Genes Chromosomes.Cancer 26:286-294.

Janin N, Andrieu N, Ossian K, Lauge A, Croquette MF, Griscelli C, Debre M, Bressac-de-Paillerets B, Aurias A, Stoppa-Lyonnet D. 1999. Breast cancer risk in ataxia telangiectasia (AT) heterozygotes: haplotype study in French AT families. Br.J.Cancer 80:1042-1045.

Knudson AG. 2001. Two genetic hits (more or less) to cancer. Nat.Rev.Cancer 1:157-162.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat.Protoc. 4:1073-1081.

Letouze E, Sow A, Petel F, Rosati R, Figueiredo BC, Burnichon N, Gimenez-Roqueplo AP, Lalli E, de RA. 2012. Identity by descent mapping of founder mutations in cancer using high-resolution tumor SNP data. PLoS.One. 7:e35897.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26:589-595.

Maxwell KN, Hart SN, Vijai J, Schrader KA, Slavin TP, Thomas T, Wubbenhorst B, Ravichandran V, Moore RM, Hu C, Guidugli L, Wenz B, Domchek SM, Robson ME, Szabo C, Neuhausen SL, Weitzel JN, Offit K, Couch FJ, Nathanson KL. 2016. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am.J.Hum.Genet. 98:801-817.

Maxwell KN, Wubbenhorst B, D'Andrea K, Garman B, Long JM, Powers J, Rathbun K, Stopfer JE, Zhu J, Bradbury AR, Simon MS, DeMichele A, Domchek SM, Nathanson KL. 2015. Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. Genet.Med. 17:630-638. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17:122.

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66-71.

Olsen JH, Hahnemann JM, Borresen-Dale AL, Tretli S, Kleinerman R, Sankila R, Hammarstrom L, Robsahm TE, Kaariainen H, Bregard A, Brondum-Nielsen K, Yuen J, Tucker M. 2005. Breast and other cancers in 1445 blood relatives of 75 Nordic patients with ataxia telangiectasia. Br.J.Cancer 93:260-265.

Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, Cheung SW, Shen RM, Barker DL, Gunderson KL. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res. 16:1136-1148.

Popova T, Manie E, Stoppa-Lyonnet D, Rigaill G, Barillot E, Stern MH. 2009. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome Biol. 10:R128.

Pylkas K, Tommiska J, Syrjakoski K, Kere J, Gatei M, Waddell N, Allinen M, Karppinen SM, Rapakko K, Kaariainen H, Aittomaki K, Blomqvist C, Mustonen A, Holli K, Khanna KK, Kallioniemi OP, Nevanlinna H, Winqvist R. 2007. Evaluation of the role of Finnish ataxiatelangiectasia mutations in hereditary predisposition to breast cancer. Carcinogenesis 28:1040-1045.

Rahman N, Stratton MR. 1998. The genetics of breast cancer susceptibility. Annu.Rev.Genet. 32:95-121.

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N. 2006. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat.Genet. 38:873-875.

Rio PG, Pernin D, Bay JO, Albuisson E, Kwiatkowski F, De LM, Bernard-Gallon DJ, Bignon YJ. 1998. Loss of heterozygosity of BRCA1, BRCA2 and ATM genes in sporadic invasive ductal breast carcinoma. Int.J.Oncol. 13:849-853.

Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy M, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu V, Goggins M, Vogelstein B, Papadopoulous N, Hruban RH, Kinzler KW, Klein AP. 2012. ATM mutations in hereditary pancreatic cancer patients. Cancer Discov. 2:41-46. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat.Biotechnol. 29:24-26.

Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. 2010. MutationTaster evaluates diseasecausing potential of sequence alterations. Nat.Methods 7:575-576.

Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, Murray A, Hanks S, Douglas J, Stratton MR, Rahman N. 2012. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res.Treat. 134:429-433.

Stratton MR, Rahman N. 2008. The emerging landscape of breast cancer susceptibility. Nat.Genet. 40:17-22.

Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de SD, Zharkikh A, Thomas A. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J.Med.Genet. 43:295-305.

Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, Byrd P, Taylor M, Easton DF. 2005. Cancer risks and mortality in heterozygous ATM mutation carriers. J.Natl.Cancer Inst. 97:813-822.

Vorechovsky I, Luo L, Lindblom A, Negrini M, Webster AD, Croce CM, Hammarstrom L. 1996. ATM mutations in cancer families. Cancer Res. 56:4130-4133.

Walsh T, King MC. 2007. Ten genes for inherited breast cancer. Cancer Cell 11:103-105.

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

FIGURE LEGENDS

Figure 1: Pedigree of the family G1995079.

The age stated below of each individual is the age at last contact or the age of death. The occurrences of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, breast cancer; LiC, liver cancer; C?, unknown primary cancer; (+), mutation carrier; (-), wild type; y, year. The figure 1B is a zoom of the pedigree showing more precisely the patients who were analyzed in our study.

Figure 2: Whole genome single nucleotide polymorphism (SNP) array profile and common loss of heterozygosity (LOH) in the two breast cancers of family G1995079.

The whole genome profile of genomic rearrangements in two breast cancers sisters was measured by HumanOmniExpress v12 Illumina SNP-array. (A) Copy-number variation profile is represented by log R ratio (LRR), centered at zero. (B) Allelic imbalances are represented by B-allele frequency (BAF). (C) The allele-specific copy number of each segment was determined by the Genome Alteration Print (GAP) method and is represented by absolute copy number. (D) Representation of LOH(s) in two familial breast cancers. A genomic segment was considered to have undergone LOH when the minor copy had a copy number of <0.5 and the LOH was classed as significant if it covered >1 Mb. Three regions of common LOH(s) were identified in the two sister's breast cancers, on chromosomes 6 [6q11.1 - 6q27], 11 [11q14.3 - 11q25] and 22 [22q11.1 - 22q13.31].

Figure 3: Detection of 7 long conserved haplotypes in the two breast cancers of family G1995079.

Visualization with Integrative Genomics Viewer (IGV) of the conserved haplotypes that were detected. (a) Representation of chromosome and region of the chromosome that is precisely concerned. (b) Tumor haplotypes are represented for both patients (III:1 and III:2) of the family G1995079, as series of blue and yellow vertical lines, corresponding to SNPs with respectively "A" and "B" genotype, according to Illumina nomenclature. The conserved haplotypes are represented by an accolade above tumor haplotypes and corresponds to the long haplotypes that are identical between the familial tumors. (c) Representation of the genes that are contained in the tumor haplotypes. Seven conserved haplotypes were detected for the both familial breast cancers: two on chromosome 6 [6q11.1 - 6q12] (A) and [6q15 - 6q27] (B), one on chromosome 11 [11q14.3 - 11q25] (C) and four on chromosome 22 [22q11.22] (D), [22q12.3] (E), [22q13.1 - 22q13.2] (F) and [22q13.31] (G). Note that only the extremities of the conserved haplotypes are represented for the figures (B) and (C).

Figure 4: Sanger sequencing of the *ATM* gene at the position of the germline *ATM* mutation identified in the family G1995079.

Sanger sequencing of leucocyte DNA (L DNA) and tumor DNA (T DNA) of patient III:1 (A) and patient III:2 (B) of family G1995079, shows a relatively low intensity of the wild-type allele in the tumors suggesting LOH of this allele at this position. The position of the *ATM* mutation (c.7570G>C – p.Ala2524Pro), is indicated with an arrow.

ADDITIONAL MATERIAL

Additional file 1: Supplementary figures.

File containing eleven supplementary figures. Sup figures 1 to 10 showing pedigrees of ten additional BC families with a germline *ATM* mutation which had been previously identified in a diagnostic setting (Sup figure 1: Pedigree of family G2010321. Sup figure 2: Pedigree of family G2014349. Sup figure 3: Pedigree of family G2012237. Sup figure 4: Pedigree of family G2012530. Sup figure 5: Pedigree of family G2011417. Sup figure 6: Pedigree of family G2009467. Sup figure 7: Pedigree of family G2014581. Sup figure 8: Pedigree of family G2013539. Sup figure 9: Pedigree of family G2014581. Sup figure 10: Pedigree of family G2014701). Sup figure 11: Sanger sequencing of germline DNA from the parents and unaffected sister of the index case of family G1995079, showing segregation of the *ATM* mutation in the family.

Additional file 2: Supplementary tables.

File containing two supplementary tables. Sup table 1: Details of the genes in the diagnostic NGS panel. Sup table 2: Primers used for amplification of the *ATM* gene.

Family_Patient		Germline mutation		Pathology	Cancer Histology
	ATM	BRCA1	BRCA2		
G1995079_III:1	c.7570G>C - p.Ala2524Pro	,	·	Left BC (30y) Right BC (32y)	IDC, grade [#] II, HR+ DCIS
G1995079_III:2	c.7570G>C - p.Ala2524Pro			BC (29y)	IDC, grade I, HR+
G1995079_II:2	wild-type			BC (68y)	IDC, grade I, HR+
G1995079_II:15				BC (51y)	IDC, grade II
G1995079_II:21				BC (71y)	IDC, grade II
G1995079_II:23				BC (50y)	IDC, grade II
G1995079_III:16				BC (NAy)	NA
BC, breast cancer; Scarff-Bloom and B	y, years old; IDC, invasive ductal (Sichardson (SBR) grade: (-). Genet	carcinoma; DCIS, ductal carcino tic analysis performed and no m	oma in situ; HR, Hormonal r nutation detected.	eceptors; NA, Not ann	otated, [#] according to
	and it then the head man minut				

Table 1. Characteristics of breast cancers in family G1995079

Family	DH_ID	TSamples	Chrom	Startpos*	Endpos*	Startcyto	Endcyto	StartSNP	EndSNP	Nsnp	Length.bp	Ngenes
	G1995079_6	III:1_III:2	9	61891118	66164344	6q11.1	6q12	rs4502981	rs12206730	566	4273226	2
	G1995079_6	III:1_III:2	9	90739172	170919470	6q15	6q27	rs3857498	rs12530134	20136	80180298	363
	G1995079_11	III:1_III:2	11	90070103	131888483	11q14.3	11q25	rs1507608	rs11525415	12139	41818380	315
G1995079	G1995079_22	III:1_III:2	22	22710758	23258369	22q11.22	22q11.22	rs17578409	rs451115	23	547611	7
	G1995079_22	III:1_III:2	22	34799261	35964722	22q12.3	22q12.3	rs5755023	rs5750116	373	1165461	9
	G1995079_22	III:1_III:2	22	39575848	41415644	22q13.1	22q13.2	rs12166894	rs4821981	319	1839796	26
	G1995079_22	111_11:2	22	44794360	47310536	22q13.31	22q13.31	rs6007363	rs4823894	951	2516176	34
LOH_ID, LOH	identity; Tsample	is, Tumor san	nples; Ch	rom, Chromo	some; Startpo	os/Startcyto/	StartSNP, Sta	art of the genor	mic/cytologic/S	NP positi	on of the con	served

Table 2. Characteristics of the 7 conserved haplotypes detected in two breast cancers of family G1995079

haplotypes; Endpos/Endcyto/EndSNP, End of the genomic/cytologic/SNP position of the conserved haplotype; Nsnp, Number of SNPs contained in each haplotype; Length.bp, length of the haplotype in base pair; Ngenes, Number of contained genes in each haplotype; *according to the hg19/ GRCh37 genome version.

lable 3. Variarits detected by	has alliova		erveu riapiorypes or	ine patient minut	DECETO AIIIIP					
Conserved Haplotype	Gene	Position (g.)*	Nucleotide (c.)	Amino acid (p.)	Type	Align GVGD	PolyPhen-2	SIFT	Mutation Taster	Splice
chr6: 61891118 - 66164344	•								.	
	UFL1	chr6:96997599	1742T>A	lle581Asn	Missense	C45 (GV: 28.68 - GD: 133.10)	Po D (0,948)	D (0)	DC (p:1)	
	BCLAF1	chr6:136599748	271G>A	Gly91Ser	Missense	C0 (GV: 109.55 - GD: 46.81)	Pr D (1)	T (0,2)	DC (p:1)	
	MAP3K5	chr6:136926510	2522-8_2522-6dup	4	Intronic					No effect
								٥		
chr6: 90739172 - 170919470	GJE1	chr6:142455821	380C>T	Ala127Val	Missense	C0 (GV: 232.67 - GD: 35.66)	B (0,006)	(0.04)	NA	
	LRP11	chr6:150147414	1334C>T	Pro445Leu	Missense	C65 (GV: 0.00 - GD: 97.78)	Pr D (1)	D (0)	DC (p:1)	
	TCP1	chr6:160201543	1029A>G	= (Ala343Ala)	Synonymous					No effect
	IGF2R	chr6:160468175	2052-16C>T	¢.	Intronic				-	No effect
	PDCD2	chr6:170886648	1034A>G	= (*345*)	Synonymous				_	No effect
								٥		
	BIRC3	chr11:102196031	791A>T	Asn264Ile	Missense	C15 (GV: 74.82 - GD: 89.28)	Po D (0,872)	(0,01)	NA	
	DYNC2H1	chr11:103026053	3574-7del	2	Intronic					No effect
	DYNC2H1	chr11:103040982	5114T>C	Leu1705Pro	Missense	C65 (GV: 4.86 - GD: 95.38)	Pr D (0,998)	D (0)	DC (p:1)	
chr11: 90070103 - 131888483	ATM	chr1:108202225	7570G>C	Ala2524Pro	Missense	C25 (GV: 0.00 - GD: 26.87)	Pr D (1)	(0) Q	DC (p:1)	
	DLAT	chr11:111896317	1216>A	Ala41Thr	Missense	C0 (GV: 353.86 - GD: 0.00)	B (0,011)	(0)(03)	Pol (p:1)	
	SIK3	chr11:116729374	2663C>T	Ala888Val	Missense	C0 (GV: 353.86 - GD: 0.00)	B(0,007)	T (0,39)	Pol (p:1)	
	AMICA1	chr11-118083270	50C>T	Ser17Phe	Missense	C25 (GV: 73.35 - GD: 113.73)	B (0.006)	D (20:0)	Pol (n:1)	
chr22: 22710758 - 23258369										
chr22: 34799261 - 35964722	ISX	chr22:35478570	289C>G	Leu97Val	Missense	C0 (GV: 234.72 - GD: 16.18)	Pr D (1)	D (0)	DC (p:1)	
chr22: 39575848 - 41415644	SGSM3	chr22:40803845	1577G>A	Arg526GIn	Missense	C0 (GV: 241.31 - GD: 0.00)	Po D (0,914)	D (0)	DC (p:1)	
chr77-44794360 - 47210536	CELSR1	chr22:46761267	8415C>T	= (Ser2805Ser)	Synonymous					No effect
	CELSR1	chr22:46795716	5310C>T	= (p.Thr1770Thr)	Synonymous					No effect
C0, C15, C25, C45 and C65 classe:	, Interference	with function from I	east (CO) to most likely	(C65); Pr D, Probably	/ damaging; Po I), Possibly damaging; B, Benign; D	, Deleterious; T,	Tolerated	; DC, Disease	causing;

upes of the patient III-1 (Family G1995079) Holney bound ancing in the Table 3. Variants detected hv Pol, Polymorphism; NA, Not annoted; p, p-value; (-) no detected variant;*according to the hg19/ GRCh37 genome version.

Comiles Potions		Germline mutation		Detholom	Canada Uistalaan
raminy_rauent	ATM	BRCA1	BRCA2	ramology	
G2010321_III:6	c.8610_8611deITA (p.Asp2870Glufs*10)	,		BC (46y)	IDC, grade [#] III, HR+, Her2-
G2010321_III:4	c.8610_8611deITA (p.Asp2870Glufs*10)		·	Right BC (35y) Left BC (47y)	IDC, grade III, HR+, Her2- IDC, grade I, HR+, Her2-
G2010321_II:4	c.8610_8611deITA (p.Asp2870Glufs*10)			BC (64y)	ILC, grade III, HR+, Her2-
G2014349_III:17	c.1236-2A>G (p.?)	,	,	BC (40y)	IDC, grade I, HR+, Her2-
G2014349_III:13	c.1236-2A>G (p.?)			BC (48y)	IDC, grade II, HR+, Her2+
G2012237_III:8	c.8873_8874delTT (p.Phe2958*)			Right BC (46y) Left BC (46y)	IDC, grade II, HR+, Her2- IDC, grade II, HR+, Her2+
G2012237_III:7	wild-type			BC (45y)	DCIS, grade I
G2012530_III:6	c.4776+1G>T (p.?)		,	BC (50y)	IDC, grade III, HR+, Her2+
G2012530_III:8	wild-type	5		BC (47y)	DCIS, grade II
G2011417_II:8	c.2098C>T (p.Glu700*)	くび	,	BC (64y)	IDC, grade II, HR+, Her2-
G2011417_III:4	wild-type			BC (44y)	IDC, grade III, HR+, Her2+
11:11 230000C			12 12 12 12 12 12 12 12 12 12 12 12 12 1	Left BC (43y) ^{T1}	ILC, grade II, HR+, Her2-
TT:11-10+60070	(DOUTLEN'S DISTORTION (DISTORT)		(T/7T IAC .d) 1077000	Right BC (43y) ⁷²	IDC, grade II, HR+, Her2-
G2011018_III:9	c.2098C>T (p.Glu700*)	•		BC (56y)	ILC, grade II, HR+, Her2-
G2013539_IV:10	c.1256T>G (p.Leu419*)		- N.	BC (45y)	IDC, grade II, HR+, Her2-
G2014581_III:2	c.3754_3756delinsCA (p.Tyr1252Glnfs*4)		- M-	BC (74y)	IDC, grade II, HR+, Her2-
G2014701_III:10	c.2113_2114insA (p.Tyr705*)	c. 4195_4196delAC (p.Thr1399Hisfs*4)		BC (31y)	IDC, grade III, HR+, Her2+
BC, breast cancer;) performed and no i	<pre>/, years old; IDC, invasive ductal carcinoma; DCIS, du mutation detected; [#] according to Scarff-Bloom and</pre>	ctal carcinoma in situ; ILC, invasive lobular carc Richardson (SBR) grade; T1 and T2, Tumor 1 ar	cinoma; HR, Hormonal receptor nd Tumor 2 of the patient G200	s; Her2, Her2Neu O 9467_III11.	ncogene; (-), Genetic analysis

Table 4. Characteristics of the 10 breast cancer families associated with germline ATM mutation

						,			
		Constitutive	SDN		Somatic NG	0		Expected mut	MC LOH
	Depth (X)	Mut allele (X)	Mut allele ratio	Depth (X)	Mut allele (X)	Mut allele ratio	cells in the sample*	allele ratio"	
G1995079_III:1			0.50			0.75 ^{\$}	0.55	0.775	Yes
G1995079_III:2			0.50 ⁵			0.75 ^{\$}	0.55	0.775	Yes
G2010321_III:6	710	341	0.48	170	144	0.85	0.70	0.85	Yes
G2010321_III:4				548	441	0.81	0.50	0.75	Yes
G2010321_II:4				110	54	0.49	0.70	0.85	No
G2014349_III:17	361	147	0.41	729	491	0.68	0.50	0.75	Yes
G2014349_III:13				921	661	0.72	0.65	0.80	Yes
G2012237_III:8	523	242	0.46	135	107	0.79	0.70	0.85	Yes
G2012530_III:6	<u> 965</u>	457	0.475	694	489	0.71	0.60	0.80	Yes
G2011417_II:8	542	275	0.51	342	222	0.65	0.70	0.85	Yes
G2009467_III:11 (T1)	N3	ų	0 10	610	391	0.64	0.40	0.70	Yes
G2009467_III:11 (T2)	\$	9	0.40	820	357	0.435	0.50	0.75	No
G2011018_III:9			0.50 ⁵	455	199	0.44	0.35	0.675	No
G2013539_IV:10	368	184	0.50	1016	721	0.71	0.55	0.775	Yes
G2014581_III:2	436	237	0.54	270	236	0.87	0.50	0.75	Yes
G2014701_III:10	473	237	0.50	1270	1143	06.0	0.80	0.90	Yes
LOH, Loss Of Heterozygo cells in the tumor sample in the analyzed samples	sity; NGS, No es; [#] Expected with the follo	ext Generation Sec d allelic ratio of thi owing formula: y =	quencing; wt, wild ty; e mutated allele (y) ii (1-x)/2 + x; ^{\$} , Evalua	pe allele; mut n the event of ted ratio from	allele, mutated alli f a loss of the wt al 1 Sanger sequencin	ele; X, read number; * lele. This ratio was cal lg: T1 and T2, Tumor :	Histological evaluation of i culated according to the pr 1 and Tumor 2 of the patier	the proportion of to oportion of tumor at G2009467_III:11	umor cells (x)

Table 5. LOH study in 16 familial breast cancers









T DNA

L DNA

∢

Gene	Chromoson	ie Genomic start*	Genomic end*	Strand orientation	of Exons	Exon size (bp)	length ¹ (bp)	consensus coding sequence ³ (bp)	UCSC Gene ID used for library capture definition	NCBI Rerseq mRNA Nomenclature ¹	Ensembl Gene Nomenclature ²	Ensembl Refere Transcript Nomenclature
4	TM 11	108 091 559	108 241 826	+	63	13147	13147	9171	uc001pkb.1	NM_000051.3	ENSG0000149311	ENST00000278
8	RD1 2	215 676 428	215 591 262		11	2607	2610	2334	uc002veu.2	NM_000465.2	ENSG00000138376	ENST0000260
8	ICA1 17	41 279 500	41 194 312		23	7207	7094	5592	uc002ict.2	NM_007294.3	ENSG0000012048	ENST0000357
8	ICA2 13	32 887 617	32 975 809	٠	27	11386	10930	10257	uc001uub.1	NM_000059.3	ENSG00000139618	ENST0000380
8	11 17	59 942 920	59 754 547	•	20	8166	6048	3750	uc002izk.1	NM_032043.2	ENSG0000136492	ENST0000259
0	DH1 16	68 769 195	68 871 444	•	16	4815	4875	2649	uc002ewg.1	NM_004360.3	ENSG0000039068	ENST0000261
0	EK2 22	29 139 822	29 081 731		14	1858	1771	1632	uc003adt.1	NM_007194.3	ENSG0000183765	ENST0000328
genes MF	E11A 11	94 229 040	94 148 466	•	20	5164	2746	2043	uc001peu.2	NM_005591.3	ENSG0000020922	ENST0000323
P	16 16	23 654 678	23 612 483		13	4058	4003	3561	uc002dlx.1	NM_024675.3	ENSG0000083093	ENST0000261
4	TEN 10	89 621 195	89 730 532	•	6	5547	5572	1212	uc001kfb.2	NM_000314.4	ENSG0000171862	ENST0000371
R	D50 5	131 890 616	131 982 313	•	25	6597	6022	3939	uc003kxi.2	NM_005732.3	ENSG0000113522	ENST000026
RA	D51C 17	56 767 963	56 813 692	٠	6	1282	1295	1131	uc002iwu.2	NM_058216.1	ENSG0000108384	ENST0000333
S	K11 19	1 203 798	1 230 434	٠	10	3276	3276	1302	uc002lrl.1	NM_000455.4	ENSG0000118046	ENST0000326
-	P53 17	7 592 868	7 569 720	•	11	2586	2579	1182	uc002gij.2	NM 000546.5	ENSG0000141510	ENST0000269
	PC 5	112 005 202	112 219 936	+	16	10730	10701	8532	uc003kpy.3	NM_000038.5	ENSG00000134982	ENST0000257
E	CAM 2	47 594 287	47 616 167	•	6	1718	1724	945	uc002rvx.2	NM_002354.2	ENSG00000119888	ENST0000263
2	LH1 3	37 032 841	37 094 337	÷	19	2662	2752	2271	uc003cgl.2	NM_000249.3	ENSG0000076242	ENST0000231
2	SH2 2	47 628 263	47 712 360	•	16	3145	3307	2805	uc002ny.1	NM_000251.2	ENSG0000095002	ENST000023
2	SH6 2	48 008 221	48 036 092	٠	10	4328	4328	4083	uc002rwd.3	NM_000179.2	ENSG0000116062	ENST000023
OTHER MI	лтүн 1	45 808 142	45 792 914		16	1921	1744	1641	uc001cnh.2	NM_012222.2	ENSG0000132781	ENST000037.
Ы	(3CA 3	178 864 311	178 954 500	•	21	3712	3724	3207	uc003fjk.2	NM_006218.2	ENSG0000121879	ENST000026:
4	VIS2 7	6 050 737	6 010 870		15	2836	2855	2589	uc003spl.2	NM_000535.5	ENSG00000122512	ENST000026
P	CH1 9	98 281 247	98 203 264		24	7943	8057	4344	uc004avk.3	NM_000264.3	ENSG00000185920	ENST0000033:
P	CH2 1	45 310 616	45 283 516		22	4298	4225	3612	uc010olf.1	NM_003738.4	ENSG00000117425	ENST0000372
S	JFU 10	104 261 719	104 395 214	+	12	4948	5001	1455	uc001kvy.1	NM_016169.3	ENSG0000107882	ENST0000369
TOTAL	25				451	125937	120386	85239				

Supp	lementary table	2: Primers used for amplificatio	n of the ATM gene						
Exon	Primer	Primer sequences (5'-> 3')	Genomic position*	Lenght	۳ (۵	ہ ⊥ ئ%	emperature of	MgCl ₂ Concentration	Size of
	ומפוומנא			(m)	5			(mm)	(bp)
ç	ATM_ex10.01F	GTCTATATATTCCTTTTAGTTTG	108121356-108121379	24	60	25	01	u t	
3	ATM_ex10.01R	CCTCTTGTCTTGACACAATGC	108121579-108121599	21	62	48	00	C'T	++7
5	ATM_ex13.02F	ACTATTICTICAGACAACTITIG	108124592-108124614	23	60	30	ĉ	u F	200
3	ATM_ex13.02R	CCCTATTTCTCCTTCCTAAC	108124822-108124841	20	28	45	0C	C'T	007
Ļ	ATM_ex15F	GTATGACTACGTGGGAACTTCT	108128050-108128070	21	60	42	ĉ	1	000
9	ATM_ex15R	CCAGTGATCAATTCCCTGCA	108128539-108128558	20	60	50	00	C'T	500
36	ATM_ex26.01F	TAATGCTGATGGTATTAAAACAG	108154855-108154877	23	60	30	ĉ	ų,	
9	ATM_ex26.01R	CTCTCTTTGCTGTGCCATCC	108155127-108155146	20	62	55	00	C'T	767
5	ATM_ex31F	ACAGTAAGTTTTGTTGGCTTAC	108163968-108163989	22	60	36	ů	5	110
7	ATM_ex31R	GCTAGAGCATTACAGATTITTG	108164260-108164281	22	60	36	00	C ⁽ T	410
5	ATM_ex51F	GCTAATAGAGGAGCACTGTC	108202071-108202090	20	60	46	U.	ų	010
10	ATM_ex51R	TITGTATTICCATTICTTAGAGG	108202367-108202389	23	60	29	00	C/T	CTC
0	ATM_ex59.01F	TATGCACATCATTTAAGTAGGC	108217865-108217886	22	60	36	6	1	216
20	ATM_ex59.01R	GTACAAGTTCTGCTGACTGC	108218061-108218080	20	60	50	0C	C'T	017
G	ATM_ex62F	TACTGGTTCTACTGTTTCTAAG	108235734-108235755	22	60	36	č	15	202
70	ATM_ex62R	AGTGAGGTGAACAGTTTAAAGG	108236005-108236026	22	62	40	00	C/1	667
* acco pair.	ording to the hg19/(5RCh37 genome version and NM_00729	4.3 transcript; Nt, Nucléotide;	°C, degree C	elsius; %	GC, GC	bercentage in prime	er; mM, milliMolaire	e; bp, base





Supplementary Figure 1: Pedigree of family G2010321 with breast cancers and germline mutation in the ATM gene.

The age stated below of each individual is the age at last contact or the age of death. The occurrences of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; BC, Breast cancer; GC, Gastric cancer; (+), mutation carrier; y, year.



The age stated below of each individual is the age at last contact or the age of death. The occurrences of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer, Half-filled symbol, other cancer, BC, Breast cancer,

PaC, Pancreatic cancer; CRC, Colorectal cancer; Lk, Leukemia; C?, unknown primary cancer; (+), mutation carrier; y, year,



G2012237



The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, Breast cancer, ThC, Throat cancer, SIDS, Sudden infant death syndrome; (+), mutation carrier; y, year;



Supplementary Figure 4: Pedigree of family G2012530 with breast cancers and germline mutation in the *ATM* gene. The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, Breast cancer; PrC, Prostate cancer; PaC, Pancreatic cancer; CRC, Colorectal cancer; C?, unknown primary cancer; (+), mutation canier; y, year;

G2012530



G2011417



Supplementary Figure 5: Pedigree of family G2011417 with breast cancers and germline mutation in the ATM gene.

The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, Breast cancer; TyC, Thyroid cancer; PaC, Pancreatic cancer; CRC, Colorectal cancer; (+), mutation carrier; y, year;



G2009467



The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; BC, Breast cancer; (+), mutation carrier; y, year;



Supplementary Figure 7: Pedigree of family G2011018 with breast cancers and germline mutation in the ATM gene.

The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, breast cancer; PrC, Prostate cancer; CRC, Colorectal cancer; Lk, Leukemia; (+), mutation carrier; y, year; The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated.

G2013539



170

The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, Breast cancer;

BraC, Brain cancer, CRC, Colorectal cancer, (+), mutation carrier, y, year,



G2014701



Supplementary Figure 10: Pedigree of family G2014701 with breast cancers and germline mutation in the *ATM* gene. The age stated beelow of each individual is the age at last contact or the age of death. The occurrence of cancer with the respective age at diagnosis are indicated. The index patients are indicated with an arrow. Striped symbol, dead person; Filled symbol, breast cancer; Half-filled symbol, other cancer; BC, Breast cancer, ThC, Throat cancer; (+), mutation carrier; y, year;

Supplementary Figure 11. Sanger sequencing of the ATM gene at the position of the germline ATM mutation identified in the family G1995079.

Sanger sequencing was performed on leucocytes DNA (L DNA) from the father (patients II:1), the mother (patient II:2) and the unaffected sister (patient III:3) of the index case of the family G1995079 to study the segregation of the ATM mutation in the family. The position of the ATM mutation (c.7570G>C - p.Ala2524Pro), is indicated with an arrow. The Sanger sequencing shows that the mutated allele was inherited from the father and that the unaffected sister was non carrier.



5 DISCUSSION ET PERSPECTIVES

5.1 L'apport des explorations tumorales dans l'identification de nouveaux gènes de prédisposition héréditaire au cancer

Les stratégies d'identification des gènes ont considérablement changé avec le développement des techniques de séquençage de nouvelle génération, le développement du NGS ayant permis de passer progressivement d'une biologie restreinte à l'étude d'un gène à une approche beaucoup plus globale.

Les évolutions spectaculaires de cette nouvelle technologie en termes de capacité et de diminution des coûts font qu'il est désormais possible d'accéder à l'intégralité du génome de la plupart des espèces et d'en étudier leur variabilité. Rappelons que le séquençage du génome humain selon la méthode de Sanger développée en 1977, avait nécessité plus de 10 ans et plusieurs milliards de dollars, pour être achevé en 2004 [409]. Aujourd'hui, le séquençage d'un exome, représentant la partie codante des gènes, soit environ 1% du génome humain, peut être réalisé en quelques jours pour moins de 2000 euros. Mais s'il est très facile de disposer aujourd'hui du séquençage d'un exome, il est beaucoup plus complexe d'interpréter les nombreuses variations génétiques identifiées et de leur attribuer une causalité, comme peuvent en témoigner les sept exomes que nous avons réalisés avec la détection d'environ 2700 INDELs et 37000 SNVs par patient. On retrouve effectivement dans la littérature, la notion d'environ 20 000 variations nucléotidiques ou SNVs détectées par exome et d'environ 500 variations rares dont la fréquence allélique est inférieure à 0.1 %, une fois les variations polymorphiques déjà répertoriées dans les bases de données éliminées [410;411]. Ces 500 variations sont susceptibles d'avoir un impact biologique même si la majorité d'entre-elles ne sont pas délétères [412]. Certaines de ces variations sont cependant privées et spécifiques à l'exome d'un individu [410].

L'une des stratégies développée pour trier ces multiples variants génétiques est l'analyse comparative d'exomes, notamment par la réalisation de trios avec soustraction à partir de l'exome du cas index, des exomes parentaux afin d'identifier des mutations *de novo*. Cette stratégie s'applique lorsqu'un seul individu est atteint dans une famille et que l'on suspecte une transmission dominante en vue d'identifier un SNV *de novo* avec un impact biologique potentiel. Vissers et al, ont décrit cette stratégie des trios d'exome chez dix patients présentant un retard mental modéré ou sévère. Six variations délétères *de novo* ont été identifiées sur des gènes reliés à des gènes du retard mental [410]. D'autres études ont ensuite confirmé l'efficacité de cette stratégie et le rôle des mutations *de novo* dans l'apparition de certaines pathologies [413;414].

Il est également possible de comparer les exomes de cas index de différentes familles atteints de la même maladie génétique pour identifier, selon la stratégie de récurrence, un gène altéré en commun. Elle recherche chez plusieurs individus sans lien de parenté, un gène commun portant une variation délétère mais elle suppose une absence d'hétérogénéité génétique. En 2010, Hoischen et ses collaborateurs ont décrit la première étude de ce type capable d'identifier un gène (*SETBP1*) responsable d'une maladie Mendélienne rare à transmission autosomique dominante (le syndrome de Schinzel-Giedion), chez quatre individus sur neuf atteints non apparentés. L'implication de ce gène a ensuite été validée chez huit individus sur neuf atteints du même syndrome par séquençage Sanger [415]. La même année, Ng et ses collaborateurs ont réalisé une étude comparative d'exomes similaire, à partir de dix patients atteints du syndrome de Kabuki mais en combinant tous les patients, les auteurs n'ont pas réussi à sélectionner le gène impliqué. L'analyse d'un nombre plus limité de patients leur a permis en revanche d'identifier le gène en cause : le gène *MLL2* [411]. Cette étude montre que cette stratégie est applicable aux pathologies ayant une homogénéité génétique et que la sélection des patients doit être restrictive pour éviter les phénocopies.

Une dernière possibilité de tri des variants est la comparaison d'exomes d'apparentés distants atteints de la même maladie génétique au sein de grandes familles, pour identifier la mutation délétère. Elle consiste à séquencer plusieurs individus de la famille, sains et malades, puis à sélectionner la/les variation(s) commune(s) aux individus atteints. Plusieurs études ont démontré que ce type d'approche intrafamiliale pouvait permettre d'identifier le gène responsable de maladies Mendéliennes monogéniques, à transmission dominante [416-418]. Mais le succès de cette stratégie repose essentiellement sur le nombre et le choix des

individus. En effet, plus les individus sont éloignés généalogiquement, moins ils partagent de matériel génétique en commun alors que le pourcentage de matériel génétique commun entre deux frères est d'environ 50 %. Ainsi, l'identification du gène responsable sera plus aisée si plusieurs individus malades sont séquencés et s'ils ne sont pas apparentés au premier degré.

Le cancer du sein, nous l'avons vu précédemment, est une pathologie rarement héréditaire (5 à 10%), se caractérisant dans ce cas par une importante hétérogénéité génétique. Il s'agit de plus d'une pathologie fréquente, touchant actuellement une femme sur huit, à l'origine d'un certain nombre de phénocopies dans les familles explorées.

Toutes ces raisons font que les différentes stratégies d'analyse comparative d'exomes précédemment décrites, ne s'appliquent pas vraiment à l'étude des formes familiales de cancer du sein pour identifier de nouveaux gènes de prédisposition à ce type de pathologie.

En revanche, la combinaison d'une approche NGS permettant l'identification de mutations privées et d'une approche tumorale par puces SNP permettant l'identification de régions candidates (ou haplotypes conservés), semblait être une stratégie d'analyse nouvelle et intéressante pour sélectionner des variants potentiellement délétères à partir d'un petit nombre d'échantillons par famille.

Letouzé et ses collaborateurs ont ainsi décrit en 2012, une stratégie d'analyse utilisant les données de puces SNP haute résolution obtenues sur des groupes de tumeurs afin d'identifier des mutations fondatrices en cause dans certains types de cancer [381]. Leur méthode était basée sur deux concepts : le premier étant que l'allèle sauvage était généralement perdu par LOH dans la tumeur des patients porteurs d'une mutation germinale (modèle de Knudson); et le second étant que la taille du fragment chromosomique hérité d'un ancêtre commun se réduisait avec les générations du fait des crossing over et conduisait à un haplotype conservé de taille minimale pour chaque patient porteur de la mutation fondatrice. Grâce à cette approche, il était donc possible de détecter des haplotypes conservés au sein de régions de LOH communes à plusieurs tumeurs et d'identifier des régions candidates à forte probabilité d'abriter des mutations germinales fondatrices dans des gènes de prédisposition au cancer. Cette méthode d'analyse a été validée par les auteurs sur deux sets de tumeurs différents : 13 cas brésiliens de corticosurrénalomes de l'enfant et 30 cas français de phéochromocytomes et paragangliomes, avec mise en évidence d'un haplotype conservé en 17p13 pour les 13 corticosurrénalomes (contenant le gène TP53 et permettant de retrouver la mutation fondatrice p.R337H) et d'un haplotype conservé en 11q23.1 pour deux des 30 phéochromocytomes et

paragangliomes (comprenant le gène *SDHD* et permettant de retrouver la mutation p.R22X). Le premier set de tumeurs constituait un groupe d'étude idéal pour valider la méthode d'analyse car extrêmement rare en l'absence de la mutation R337H. Le deuxième set de tumeurs montrait en revanche que la méthode développée était également capable de détecter une mutation fondatrice présente dans seulement deux échantillons.

Letouzé et ses collaborateurs ont ainsi démontré que leur approche était tout à fait adaptée à l'identification de gènes de prédisposition au cancer. Nous avons alors choisi de l'appliquer à des familles BRCAx dans l'idée d'identifier des haplotypes communs dans les aberrations chromosomiques propres aux tumeurs ayant touché différents membres d'une même famille puisque seulement deux échantillons tumoraux par famille suffisaient pour l'entreprendre. L'objectif ici n'était pas d'identifier une mutation fondatrice commune à un groupe de cancers du sein mais des mutations privées, propres à chaque famille. Ces mutations ont été recherchées secondairement par NGS, dans les régions candidates identifiées.

Cette approche ne s'est malheureusement pas révélée très concluante puisque pour la famille témoin d'une mutation *BRCA2* (G2009373), l'une des deux tumeurs familiales ne présentait pas de LOH au locus de *BRCA2*, ne permettant pas dans ce cas l'identification du gène *BRCA2* comme gène responsable des cancers du sein observés dans cette famille. Cet exemple souligne les limites de l'approche, nécessitant que l'allèle sauvage du gène en cause soit inactivé par LOH dans la tumeur. De plus, concernant les autres familles étudiées, cette approche n'a pas permis l'identification de nouveaux gènes de prédisposition au cancer du sein. En revanche, elle a permis d'identifier *ATM* comme gène responsable d'une récurrence familiale de cancer du sein longtemps restée inexpliquée par l'exploration de plusieurs membres de la famille et selon différentes techniques d'analyse.

Cette approche reste malgré tout intéressante car elle permet le tri des nombreux variants génétiques détectés par NGS sur l'hypothèse de l'inactivation d'un gène suppresseur de tumeur répondant au modèle de Knudson, concept selon lequel l'oncogénétique s'est majoritairement fondée. Elle nécessite de plus peu d'échantillons (≥ 2) par famille, permettant de multiplier les familles à la recherche de mutations privées. Elle impose en revanche des échantillons tumoraux de bonne qualité. En effet, il n'a pas été possible de génotyper les tumeurs de six familles du fait d'une contamination des échantillons par des cellules non tumorales, réduisant ainsi le nombre de familles étudiées à 11. L'exploration de familles BRCAx supplémentaires paraît nécessaire pour mieux juger de la pertinence de cette
approche. Aussi, il semblerait intéressant d'étendre notre stratégie aux régions communes de duplication ou d'amplification, dans l'hypothèse inverse de l'implication d'un oncogène.

5.2 L'implication du gène *ATM* dans la prédisposition héréditaire au cancer du sein

Les exomes réalisés dans les régions candidates identifiées par génotypage tumoral de sept familles BRCAx ont permis d'identifier une dizaine de variants faux sens pouvant être intéressants dont un concernant le gène *ATM*. Bien qu'il s'agisse d'un variant faux sens, la mutation *ATM* identifiée dans la famille G1995079 était clairement considérée délétère sur le plan bibliographique car rapportée dans des familles AT à l'état homozygote ou en association avec une mutation *frameshift* avec un excès de cas de cancer du sein observé chez les hétérozygotes porteurs de la mutation, ainsi que dans des formes familiales de cancer du sein sans contexte d'AT associé [405;406]. L'ensemble de ces observations plaidait en faveur de sa pathogénicité et de son implication dans le risque tumoral mammaire, ce qui n'était pas le cas des dix autres variants faux sens simplement prédits comme potentiellement délétères selon des logiciels de modélisation protéique.

C'est de plus grâce à l'identification de cette mutation *ATM*, au sein d'une famille restée longtemps inexpliquée sur le plan moléculaire malgré de nombreuses explorations génétiques entreprises chez différents membres, que nous nous sommes repenchés sur le rôle que pouvait avoir le gène *ATM* dans la prédisposition héréditaire au cancer du sein et que nous avons élargi notre étude à dix familles supplémentaires. En effet, ce gène tout d'abord connu pour être responsable à l'état homozygote du syndrome d'Ataxie-télangiectasie, une affection prédisposant notamment au cancer de l'enfant, a ensuite été mis en rapport avec un risque augmenté de cancer du sein chez les apparentées hétérozygotes d'enfants AT puis, plus tardivement, chez les hétérozygotes en dehors de tout contexte familial d'AT. Depuis 2006, le risque de cancer du sein lié aux mutations hétérozygotes du gène *ATM* est connu mais étant modéré, il était jusqu'à présent considéré comme équivalent au risque d'une personne ayant

un antécédent familial de cancer du sein au premier degré, ne justifiant pas de surveillance différente que celle préconisée dans ce cas.

Plus récemment, les études de familles BRCAx par des approches NGS utilisant des panels de gènes impliqués, ou potentiellement impliqués, dans la prédisposition héréditaire au cancer du sein ont montré que les mutations constitutionnelles du gène ATM étaient retrouvées avec une fréquence bien plus élevée que dans la population générale (2-5% [419-423] contre 0,5-1% [174;306]), suggérant une surreprésentation des mutations ATM dans le groupe des familles BRCAx. Par une approche similaire, Maxwell et ses collaborateurs, ont réalisé des exomes sur 389 individus appartenant à 243 familles considérées à haut risque de cancer du sein et/ou de l'ovaire avec mise en évidence d'une mutation ATM dans 2,9% des cas. De manière intéressante, les mutations concernant le gène ATM étaient significativement plus fréquentes dans le groupe des individus avec des formes familiales ou des formes précoces de cancer du sein que dans le groupe avec des atteintes multiples ou des cancers de l'ovaire [424]. Ces données de la littérature sont finalement assez concordantes avec nos propres résultats puisque les 11 familles de l'article que nous avons étudiées dans ce contexte, se caractérisaient toutes par une agrégation familiale de cancer du sein avec un âge moyen au diagnostic de 45 ans [min 29 ans – max 74 ans] pour les 19 tumeurs associées à une mutation constitutionnelle du gène ATM, et aucun cas de cancer de l'ovaire n'était retrouvé parmi elles. En revanche, nous avons pu remarquer la survenue de cas de cancer du pancréas (au moins trois) ainsi que des atteintes bilatérales (six cas parmi 4/11 familles) comme déjà rapportés chez les hétérozygotes ATM [425;426].

Pour démontrer l'implication d'*ATM* dans la carcinogénèse mammaire, nous nous sommes intéressés aux pertes d'hétérozygotie touchant celui-ci dans 16 tumeurs du sein associées à une mutation constitutionnelle du gène *ATM*. La perte d'hétérozygotie au locus du gène *ATM* est cependant un évènement fréquent, rapporté dans 40% des cancers du sein sporadiques [312;427]. Si l'on suppose qu'*ATM* n'est pas la cible spécifique de cette délétion, cet évènement génomique peut concerner l'allèle sauvage comme l'allèle muté, chacun avec une fréquence de 20%. Or dans notre étude, nous avons trouvé que plus de 80% (13/16; p < 0.001) des tumeurs étudiées avaient perdu l'allèle sauvage.

Par ailleurs, deux familles parmi les 11 étudiées présentaient en plus de la mutation ATM identifiée, soit une mutation BRCA1 (famille G2014701) soit une mutation BRCA2

(G2009467). A notre connaissance, il n'a pas été rapporté un plus haut risque tumoral chez les porteurs hétérozygotes d'une double mutation *BRCA* et *ATM*. Concernant la tumeur T1 de la patiente G2009467_III :11, nous avons observé une perte de l'allèle sauvage autant pour le gène *ATM* que pour le gène *BRCA2*, alors que pour la tumeur T2 de la même patiente, nous n'avons observé aucune LOH pour les deux gènes. Concernant la tumeur de la patiente G2014701_III :10, les ratios alléliques montraient non seulement une perte de l'allèle sauvage pour *ATM* mais également une perte de l'allèle muté pour *BRCA1* suggérant que la survenue de cette tumeur n'était pas liée à la mutation *BRCA1* mais associée à la mutation *ATM*. L'ensemble de ces résultats plaide fortement en faveur de l'implication du gène *ATM* dans la carcinogénèse de ces 13 tumeurs, tel un gène suppresseur, et indique qu'il est un gène de susceptibilité au cancer du sein impliqué dans quelques formes familiales.

En revanche, le bilan que nous pouvons faire des études de coségrégation des mutations *ATM* identifiées avec les atteintes familiales rapportées dans certaines familles est plus mitigé dans la mesure où elle n'est observée que dans trois familles sur les six informatives. Ceci est cependant en accord avec les résultats d'études précédentes montrant la faible coségrégation et le risque modéré de cancer du sein en cas de mutations *ATM* [174;407]. Toutefois, dans notre étude comme dans la littérature, le nombre de familles explorées est assez restreint (\leq 6). L'étude d'un plus grand nombre de familles *ATM* avec des apparentés atteints disponibles, est nécessaire pour espérer mettre en évidence une quelconque différence significative.

L'ensemble de ces nouvelles données montrant que le gène *ATM* est fortement impliqué dans la cancérogénèse mammaire de formes familiales de cancers du sein, renforce les données d'épidémiologie génétique publiées à son sujet et permet de le resituer parmi les gènes de prédisposition héréditaire au cancer du sein. Ceci soulève cependant la question de son criblage moléculaire dans le cadre du diagnostic génétique des formes héréditaires de cancer du sein au même titre que les gènes *BRCA1/2* et *PALB2* analysés actuellement, d'autant que la connaissance d'une mutation *ATM* pourrait avoir un intérêt thérapeutique. En effet, il été rapporté que les femmes porteuses d'une mutation *ATM* délétère, prise en charge pour un cancer du sein controlatéral que les femmes porteuses d'une mutation similaire non irradiées [428]. Il est de plus entendu que les tumeurs présentant un défaut de réparation des cassures double-brin de l'ADN par recombinaison homologue, dont les tumeurs ATM font partie, sont plus sensibles à certaines molécules comme les sels de platine et les

inhibiteurs de PARP [313]. Enfin, le diagnostic moléculaire des mutations *ATM* permettrait également de proposer aux femmes concernées une surveillance mammaire plus adaptée, au moins annuelle et à partir de l'âge de 30 ans au vu des âges de survenue rapportés dans notre étude.

5.3 L'hérédité manquante dans les formes familiales de cancer du sein

Ce travail, comme d'autres auparavant, montre que les gènes majeurs de prédisposition héréditaire au cancer du sein ont déjà été identifiés. Les familles BRCAx sont donc liées à des mutations dans d'autres gènes, de pénétrance modérée, tels que le gène *ATM*, pour lesquels les niveaux de risque tumoral et l'expression clinique de la maladie sont très probablement dépendant de l'action de gènes modificateurs.

Dans la littérature, il est effectivement retrouvé des exemples concrets de gènes modificateurs dans des syndromes de prédisposition au cancer connu, comme en témoigne la survenue plus précoce de tumeurs (5 ans avant l'âge moyen) chez des patients Li-Fraumeni lorsque la mutation du gène en cause, *TP53*, est associée à l'haplotype 285G–309G (pour c.14+309T>G ; c.14+285G>C) du gène *MDM2* considéré comme gène modificateur [429]. L'existence d'un digénisme a également pu être rapportée dans une forme héréditaire de cancer colorectal, par la mise en évidence d'une mutation d'épissage touchant l'exon 1 du gène *OGG1* et d'une mutation faux sens du gène *MUTYH* chez une patiente décédée à 37 ans de l'évolution d'un cancer colorectal diagnostiqué un an plus tôt. Les apparentés du cas index, porteurs de l'une des deux mutations, présentaient quant à eux un phénotype moins sévère, à type d'adénomes ou de cancer colorectal après l'âge de 55 ans [430].

De manière intéressante, la mutation *ATM* faux sens identifiée dans la famille G1995079, qui ségrége avec les atteintes précoces développées par deux sœurs, a été héritée de leur père et non de leur mère ayant elle-même présentée un cancer du sein. L'absence de cas de cancer du sein dans la lignée paternelle peut s'expliquer par une généalogie peu étendue et constituée à

majorité d'hommes. La lignée maternelle qui n'est pas concernée par cette mutation, montre en revanche une agrégation notable de cancers du sein avec cinq cas rapportés, respectivement chez la mère, trois tantes et une cousine germaine maternelles. Toutes ces atteintes sont survenues tardivement, entre 50 et 71 ans, contrastant avec les pathologies présentées par les deux sœurs, entre 29 et 32 ans, dont elles sont décédées rapidement malgré un pronostic initial favorable (tumeurs de type luminale A) et une prise en charge optimale. Cette observation pose effectivement la question d'un digénisme pour expliquer les formes précoces et sévères des cancers du sein observés chez les deux sœurs, la mutation ATM paternelle dont elles sont porteuses pouvant dans ce cas être associée à un allèle de susceptibilité d'origine maternelle non identifié à l'heure actuelle. Des études supplémentaires sont nécessaires pour explorer cette hypothèse d'autant que si l'on reprend les résultats d'exome du cas index de cette famille (patiente G1995079 III:1), on retrouve dans les régions candidates identifiées par génotypage tumoral, deux variants faux sens intéressants, concernant respectivement les gènes UFL1 (sur le chromosome 6) et SGSM3 (sur le chromosome 22), décrits dans la littérature comme potentiels gènes suppresseurs de tumeur ou agissant étroitement avec des gènes suppresseurs de tumeurs connus [399;403;404]. Il a d'ailleurs été décrit récemment que l'ufmylation d'ASC1 (pour Activating signal cointegrator 1) par le système UFM1 (pour *Ubiquitin-fold modifier 1*) dont fait partie l'E3 ligase UFL1, jouerait un rôle dans le développement du cancer du sein en permettant la transactivation du récepteur aux estrogènes REa [431]. L'exploration d'un phénomène de digénisme peut permettre de mieux comprendre le cas précis de cette famille mais peut certainement s'étendre à de très nombreuses autres situations cliniques de formes familiales de cancer de sein pour lesquelles le diagnostic moléculaire ne permet pas à l'heure actuelle d'identifier de façon formelle un syndrome de prédisposition pourtant fortement pressenti.

6 RÉFÉRENCES BIBLIOGRAPHIQUES

- (1) Baselga J, Norton L. Focus on breast cancer. Cancer Cell 2002 May;1:319-22.
- (2) INCa. Les cancers en France, Les Données, INCa, édition 2015. Boulogne-Billancourt: Institut national du Cancer (INCa); 2016.
- (3) Remontet L, Buemi A, Velten M, Jougla E, Estève J. Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Institut de veille sanitaire, Saint-Maurice. InVS; Réseau Francim; Inserm; 2003.
- (4) Remontet L, Esteve J, Bouvier AM, et al. Cancer incidence and mortality in France over the period 1978-2000. Rev Epidemiol Sante Publique 2003 Feb;51:3-30.
- (5) Molinie F, Vanier A, Woronoff AS, et al. Trends in breast cancer incidence and mortality in France 1990-2008. Breast Cancer Res Treat 2014 Aug;147:167-75.
- (6) Glass AG, Lacey JV, Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst 2007 Aug 1;99:1152-61.
- (7) Allemand H, Seradour B, Weill A, Ricordeau P. [Decline in breast cancer incidence in 2005 and 2006 in France: a paradoxical trend]. Bull Cancer 2008 Jan;95:11-5.
- (8) Cancer et environnement -Inserm Expertise collective. AFFSSET; 2008.
- (9) Kelsey JL, Bernstein L. Epidemiology and prevention of breast cancer. Annu Rev Public Health 1996;17:47-67.
- (10) McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancerepidemiology, risk factors, and genetics. BMJ 2000 Sep 9;321:624-8.
- (11) Hill C. [Breast cancer screening after age 70]. Bull Cancer 2003 Dec;90:1035-7.
- (12) Friedenreich CM, Orenstein MR. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. J Nutr 2002 Nov;132:3456S-64S.
- (13) Tehard B, Friedenreich CM, Oppert JM, Clavel-Chapelon F. Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. Cancer Epidemiol Biomarkers Prev 2006 Jan;15:57-64.
- (14) Longnecker MP. Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review. Cancer Causes Control 1994 Jan;5:73-82.
- (15) Smith-Warner SA, Spiegelman D, Yaun SS, et al. Alcohol and breast cancer in women: a pooled analysis of cohort studies. JAMA 1998 Feb 18;279:535-40.

- (16) Hamajima N, Hirose K, Tajima K, et al. Alcohol, tobacco and breast cancer-collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 2002 Nov 18;87:1234-45.
- (17) Ellison RC, Zhang Y, McLennan CE, Rothman KJ. Exploring the relation of alcohol consumption to risk of breast cancer. Am J Epidemiol 2001 Oct 15;154:740-7.
- (18) Patterson RE, Cadmus LA, Emond JA, Pierce JP. Physical activity, diet, adiposity and female breast cancer prognosis: a review of the epidemiologic literature. Maturitas 2010 May;66:5-15.
- (19) Soldin OP, Makambi KH, Soldin SJ, O'Mara DM. Steroid hormone levels associated with passive and active smoking. Steroids 2011 Jun;76:653-9.
- Brunet JS, Ghadirian P, Rebbeck TR, et al. Effect of smoking on breast cancer in carriers of mutant BRCA1 or BRCA2 genes. J Natl Cancer Inst 1998 May 20;90:761-6.
- (21) Xue F, Willett WC, Rosner BA, Hankinson SE, Michels KB. Cigarette smoking and the incidence of breast cancer. Arch Intern Med 2011 Jan 24;171:125-33.
- (22) Rudel RA, Attfield KR, Schifano JN, Brody JG. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. Cancer 2007 Jun 15;109:2635-66.
- (23) Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. Lancet Oncol 2001 Mar;2:133-40.
- (24) Cuzick J. Epidemiology of breast cancer--selected highlights. Breast 2003 Dec;12:405-11.
- (25) Boyd NF, Dite GS, Stone J, et al. Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med 2002 Sep 19;347:886-94.
- (26) Martin LJ, Melnichouk O, Guo H, et al. Family history, mammographic density, and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2010 Feb;19:456-63.
- (27) McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancerepidemiology, risk factors, and genetics. BMJ 2000 Sep 9;321:624-8.
- (28) Boice JD, Jr. Cancer following irradiation in childhood and adolescence. Med Pediatr Oncol Suppl 1996;1:29-34.
- (29) Little MP, Muirhead CR, Haylock RG, Thomas JM. Relative risks of radiationassociated cancer: comparison of second cancer in therapeutically irradiated populations with the Japanese atomic bomb survivors. Radiat Environ Biophys 1999 Dec;38:267-83.
- (30) Doody MM, Lonstein JE, Stovall M, et al. Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine (Phila Pa 1976) 2000 Aug 15;25:2052-63.

- (31) Hoffman DA, Lonstein JE, Morin MM, et al. Breast cancer in women with scoliosis exposed to multiple diagnostic x rays. J Natl Cancer Inst 1989 Sep 6;81:1307-12.
- (32) Alm El-Din MA, El-Badawy SA, Taghian AG. Breast cancer after treatment of Hodgkin's lymphoma: general review. Int J Radiat Oncol Biol Phys 2008 Dec 1;72:1291-7.
- (33) Brillaud V, Tunon de LC, Richaud P, et al. [Hodgkin's disease and breast cancer: for which patients is there an association? A series from Institut Bergonie]. Gynecol Obstet Fertil 2012 May;40:284-90.
- (34) Kelsey JL, Horn-Ross PL. Breast cancer: magnitude of the problem and descriptive epidemiology. Epidemiol Rev 1993;15:7-16.
- (35) Clavel-Chapelon F, Gerber M. Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? Breast Cancer Res Treat 2002 Mar;72:107-15.
- (36) Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002 Apr 17;94:606-16.
- (37) Key TJ, Appleby PN, Reeves GK, et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol 2013 Sep;14:1009-19.
- (38) Zeleniuch-Jacquotte A, Shore RE, Koenig KL, et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. Br J Cancer 2004 Jan 12;90:153-9.
- (39) Breast cancer and hormonal contraceptives: further results. Collaborative Group on Hormonal Factors in Breast Cancer. Contraception 1996 Sep;54:1S-106S.
- (40) Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 1997 Oct 11;350:1047-59.
- (41) Fournier A, Berrino F, Clavel-Chapelon F. Unequal risks for breast cancer associated with different hormone replacement therapies: results from the E3N cohort study. Breast Cancer Res Treat 2008 Jan;107:103-11.
- (42) Hulka BS, Moorman PG. Breast cancer: hormones and other risk factors. Maturitas 2001 Feb 28;38:103-13.
- (43) Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet 2002 Jul 20;360:187-95.
- (44) Lyons TR, Schedin PJ, Borges VF. Pregnancy and breast cancer: when they collide. J Mammary Gland Biol Neoplasia 2009 Jun;14:87-98.

- (45) Beral V, Bull D, Doll R, Peto R, Reeves G. Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83?000 women with breast cancer from 16 countries. Lancet 2004 Mar 27;363:1007-16.
- (46) Kirschner MA, Samojlik E, Drejka M, et al. Androgen-estrogen metabolism in women with upper body versus lower body obesity. J Clin Endocrinol Metab 1990 Feb;70:473-9.
- (47) Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med 2006 Jan 19;354:270-82.
- (48) Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001 Oct 27;358:1389-99.
- (49) Lalloo F, Evans DG. Familial breast cancer. Clin Genet 2012 Aug;82:105-14.
- (50) Economopoulou P, Dimitriadis G, Psyrri A. Beyond BRCA: new hereditary breast cancer susceptibility genes. Cancer Treat Rev 2015 Jan;41:1-8.
- (51) Carroll JC, Cremin C, Allanson J, et al. Hereditary breast and ovarian cancers. Can Fam Physician 2008 Dec;54:1691-2.
- (52) Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 1991 Feb;48:232-42.
- (53) Lakhani SR. WHO classification of tumours of the breast. Lyon: International Agency for Research on Cancer; 2012.
- (54) Ridolfi RL, Rosen PP, Port A, Kinne D, Mike V. Medullary carcinoma of the breast: a clinicopathologic study with 10 year follow-up. Cancer 1977 Oct;40:1365-85.
- (55) Dabakuyo TS, Bonnetain F, Roignot P, et al. Population-based study of breast cancer survival in Cote d'Or (France): prognostic factors and relative survival. Ann Oncol 2008 Feb;19:276-83.
- (56) BLOOM HJ, RICHARDSON WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957 Sep;11:359-77.
- (57) Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991 Nov;19:403-10.
- (58) Blamey RW, Ellis IO, Pinder SE, et al. Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. Eur J Cancer 2007 Jul;43:1548-55.
- (59) Blamey RW, Pinder SE, Ball GR, et al. Reading the prognosis of the individual with breast cancer. Eur J Cancer 2007 Jul;43:1545-7.

- (60) Reiner A, Reiner G, Spona J, Schemper M, Holzner JH. Histopathologic characterization of human breast cancer in correlation with estrogen receptor status. A comparison of immunocytochemical and biochemical analysis. Cancer 1988 Mar 15;61:1149-54.
- (61) Macgrogan G, de M, I, Sierankowski G, et al. Time for reappraisal of progesteronereceptor testing in breast cancer management. J Clin Oncol 2005 Apr 20;23:2870-1.
- (62) Tal M, Wetzler M, Josefberg Z, et al. Sporadic amplification of the HER2/neu protooncogene in adenocarcinomas of various tissues. Cancer Res 1988 Mar 15;48:1517-20.
- (63) McNeil C. Herceptin raises its sights beyond advanced breast cancer. J Natl Cancer Inst 1998 Jun 17;90:882-3.
- (64) Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 2005 Oct 20;353:1659-72.
- (65) Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005 Oct 20;353:1673-84.
- (66) Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000 Aug 17;406:747-52.
- (67) Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001 Sep 11;98:10869-74.
- (68) Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003 Jul 8;100:8418-23.
- (69) Sorlie T. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. Eur J Cancer 2004 Dec;40:2667-75.
- (70) Comprehensive molecular portraits of human breast tumours. Nature 2012 Oct 4;490:61-70.
- (71) Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics 2006;7:127.
- (72) Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
- (73) Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? Mol Oncol 2010 Jun;4:192-208.
- (74) Abd El-Rehim DM, Pinder SE, Paish CE, et al. Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol 2004 Jun;203:661-71.

- (75) Wang XX, Fu L, Li X, et al. Somatic mutations of the mixed-lineage leukemia 3 (MLL3) gene in primary breast cancers. Pathol Oncol Res 2011 Jun;17:429-33.
- (76) Geyer FC, Lopez-Garcia MA, Lambros MB, Reis-Filho JS. Genetic characterization of breast cancer and implications for clinical management. J Cell Mol Med 2009 Oct;13:4090-103.
- (77) Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003 Sep 2;100:10393-8.
- (78) Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 2005 Jul 7;24:4660-71.
- (79) Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
- (80) Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium. Lancet 1997 May 24;349:1505-10.
- (81) Eisinger F, Jacquemier J, Charpin C, et al. Mutations at BRCA1: the medullary breast carcinoma revisited. Cancer Res 1998 Apr 15;58:1588-92.
- (82) Jacquemier J, Eisinger F, Nogues C, et al. Histological type and syncytial growth pattern affect E-cadherin expression in a multifactorial analysis of a combined panel of sporadic and BRCA1-associated breast cancers. Int J Cancer 1999 Sep 24;83:45-9.
- (83) Kleer CG. Carcinoma of the breast with medullary-like features: diagnostic challenges and relationship with BRCA1 and EZH2 functions. Arch Pathol Lab Med 2009 Nov;133:1822-5.
- (84) van der Groep P, van Diest PJ, Menko FH, et al. Molecular profile of ductal carcinoma in situ of the breast in BRCA1 and BRCA2 germline mutation carriers. J Clin Pathol 2009 Oct;62:926-30.
- (85) Eisinger F, Stoppa-Lyonnet D, Longy M, et al. Germ line mutation at BRCA1 affects the histoprognostic grade in hereditary breast cancer. Cancer Res 1996 Feb 1;56:471-4.
- (86) Johannsson OT, Idvall I, Anderson C, et al. Tumour biological features of BRCA1induced breast and ovarian cancer. Eur J Cancer 1997 Mar;33:362-71.
- (87) Armes JE, Egan AJ, Southey MC, et al. The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. Cancer 1998 Dec 1;83:2335-45.
- (88) Atchley DP, Albarracin CT, Lopez A, et al. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol 2008 Sep 10;26:4282-8.

- (89) Lakhani SR, van de Vijver MJ, Jacquemier J, et al. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. J Clin Oncol 2002 May 1;20:2310-8.
- (90) Hagen AI, Bofin AM, Ytterhus B, et al. Amplification of TOP2A and HER-2 genes in breast cancers occurring in patients harbouring BRCA1 germline mutations. Acta Oncol 2007;46:199-203.
- (91) Foulkes WD, Stefansson IM, Chappuis PO, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003 Oct 1;95:1482-5.
- (92) Armes JE, Trute L, White D, et al. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. Cancer Res 1999 Apr 15;59:2011-7.
- (93) Aaltonen K, Blomqvist C, Amini RM, et al. Familial breast cancers without mutations in BRCA1 or BRCA2 have low cyclin E and high cyclin D1 in contrast to cancers in BRCA mutation carriers. Clin Cancer Res 2008 Apr 1;14:1976-83.
- (94) Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001 Feb 22;344:539-48.
- (95) van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002 Jan 31;415:530-6.
- (96) Palacios J, Honrado E, Osorio A, et al. Re: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2004 May 5;96:712-4.
- (97) van der Groep P, Bouter A, van der Zanden R, et al. Re: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2004 May 5;96:712-3.
- (98) Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol 2007;8:R157.
- (99) Fernandez-Ramires R, Sole X, De CL, et al. Gene expression profiling integrated into network modelling reveals heterogeneity in the mechanisms of BRCA1 tumorigenesis. Br J Cancer 2009 Oct 20;101:1469-80.
- (100) Yan M, Rayoo M, Takano EA, Fox SB. BRCA1 tumours correlate with a HIF-1alpha phenotype and have a poor prognosis through modulation of hydroxylase enzyme profile expression. Br J Cancer 2009 Oct 6;101:1168-74.
- (101) Tung N, Wang Y, Collins LC, et al. Estrogen receptor positive breast cancers in BRCA1 mutation carriers: clinical risk factors and pathologic features. Breast Cancer Res 2010;12:R12.
- (102) Tirkkonen M, Johannsson O, Agnarsson BA, et al. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. Cancer Res 1997 Apr 1;57:1222-7.

- (103) Wessels LF, van WT, Hart AA, et al. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. Cancer Res 2002 Dec 1;62:7110-7.
- (104) van Beers EH, van WT, Wessels LF, et al. Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations. Cancer Res 2005 Feb 1;65:822-7.
- (105) Joosse SA, van Beers EH, Tielen IH, et al. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. Breast Cancer Res Treat 2009 Aug;116:479-89.
- (106) Price GR, Armes JE, Ramus SJ, et al. Phenotype-directed analysis of genotype in early-onset, familial breast cancers. Pathology 2006 Dec;38:520-7.
- (107) Melchor L, Honrado E, Garcia MJ, et al. Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes. Oncogene 2008 May 15;27:3165-75.
- (108) Stefansson OA, Jonasson JG, Johannsson OT, et al. Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. Breast Cancer Res 2009;11:R47.
- (109) Waddell N, Arnold J, Cocciardi S, et al. Subtypes of familial breast tumours revealed by expression and copy number profiling. Breast Cancer Res Treat 2010 Oct;123:661-77.
- (110) Vincent-Salomon A, Macgrogan G, Charaffe-Jauffret E, Jacquemier J, Arnould L. [Identification of basal-like carcinomas in clinical practice: "triple zero/BRCA1-like" carcinomas]. Bull Cancer 2010 Mar;97:357-63.
- (111) Honrado E, Osorio A, Palacios J, Benitez J. Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations. Oncogene 2006 Sep 25;25:5837-45.
- (112) Grushko TA, Dignam JJ, Das S, et al. MYC is amplified in BRCA1-associated breast cancers. Clin Cancer Res 2004 Jan 15;10:499-507.
- (113) Kauraniemi P, Hedenfalk I, Persson K, et al. MYB oncogene amplification in hereditary BRCA1 breast cancer. Cancer Res 2000 Oct 1;60:5323-8.
- (114) Staff S, Nupponen NN, Borg A, Isola JJ, Tanner MM. Multiple copies of mutant BRCA1 and BRCA2 alleles in breast tumors from germ-line mutation carriers. Genes Chromosomes Cancer 2000 Aug;28:432-42.
- (115) Popova T, Manie E, Stoppa-Lyonnet D, et al. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. Genome Biol 2009;10:R128.
- (116) Popova T, Manie E, Rieunier G, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. Cancer Res 2012 Nov 1;72:5454-62.

- (117) Manie E, Popova T, Battistella A, et al. Genomic hallmarks of homologous recombination deficiency in invasive breast carcinomas. Int J Cancer 2016 Feb 15;138:891-900.
- (118) Sobol H, Stoppalyonnet D, Bressacdepaillerets B, et al. BRCA1-p53 relationship in hereditary breast cancer. Int J Oncol 1997 Feb;10:349-53.
- (119) Greenblatt MS, Chappuis PO, Bond JP, Hamel N, Foulkes WD. TP53 mutations in breast cancer associated with BRCA1 or BRCA2 germ-line mutations: distinctive spectrum and structural distribution. Cancer Res 2001 May 15;61:4092-7.
- (120) Foulkes WD, Chappuis PO, Wong N, et al. Primary node negative breast cancer in BRCA1 mutation carriers has a poor outcome. Ann Oncol 2000 Mar;11:307-13.
- (121) Crook T, Crossland S, Crompton MR, Osin P, Gusterson BA. p53 mutations in BRCA1-associated familial breast cancer. Lancet 1997 Aug 30;350:638-9.
- (122) Saal LH, Gruvberger-Saal SK, Persson C, et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. Nat Genet 2008 Jan;40:102-7.
- (123) Lakhani SR, Jacquemier J, Sloane JP, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J Natl Cancer Inst 1998 Aug 5;90:1138-45.
- (124) Bane AL, Beck JC, Bleiweiss I, et al. BRCA2 mutation-associated breast cancers exhibit a distinguishing phenotype based on morphology and molecular profiles from tissue microarrays. Am J Surg Pathol 2007 Jan;31:121-8.
- (125) Chen J, Silver DP, Walpita D, et al. Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells. Mol Cell 1998 Sep;2:317-28.
- (126) Colombo M, Giarola M, Mariani L, et al. Cyclin D1 expression analysis in familial breast cancers may discriminate BRCAX from BRCA2-linked cases. Mod Pathol 2008 Oct;21:1262-70.
- (127) Palacios J, Honrado E, Osorio A, et al. Immunohistochemical characteristics defined by tissue microarray of hereditary breast cancer not attributable to BRCA1 or BRCA2 mutations: differences from breast carcinomas arising in BRCA1 and BRCA2 mutation carriers. Clin Cancer Res 2003 Sep 1;9:3606-14.
- (128) Rouault A, Banneau G, Macgrogan G, et al. Deletion of chromosomes 13q and 14q is a common feature of tumors with BRCA2 mutations. PLoS One 2012;7:e52079.
- (129) Jonsson G, Naylor TL, Vallon-Christersson J, et al. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 2005 Sep 1;65:7612-21.
- (130) Jonsson G, Staaf J, Vallon-Christersson J, et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res 2010;12:R42.

- (131) van Beers EH, Nederlof PM. Array-CGH and breast cancer. Breast Cancer Res 2006;8:210.
- (132) Joosse SA, Brandwijk KI, Devilee P, et al. Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. Breast Cancer Res Treat 2012 Apr;132:379-89.
- (133) Pecuchet N, Popova T, Manie E, et al. Loss of heterozygosity at 13q13 and 14q32 predicts BRCA2 inactivation in luminal breast carcinomas. Int J Cancer 2013 Dec 15;133:2834-42.
- (134) Le Dran H. Mémoire avec un précis de plusieurs observations sur le cancer. Mémoires de l'Académie Royale de Chirurgie; 1757.
- (135) Eisinger F, Sobol H, Serin D, Whorton JC. Hereditary breast cancer, circa 1750. Lancet 1998 May 2;351:1366.
- (136) Broca P. Traité des tumeurs. Asselin P.; 1866.
- (137) Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971 Apr;68:820-3.
- (138) Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000 Jul 13;343:78-85.
- (139) Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 2000 Dec;26:411-4.
- (140) Iselius L, Slack J, Littler M, MORTON NE. Genetic epidemiology of breast cancer in Britain. Ann Hum Genet 1991 May;55:151-9.
- (141) Williams WR, Anderson DE. Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1984;1:7-20.
- (142) Bishop DT, Cannon-Albright L, McLellan T, Gardner EJ, Skolnick MH. Segregation and linkage analysis of nine Utah breast cancer pedigrees. Genet Epidemiol 1988;5:151-69.
- (143) Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990 Dec 21;250:1684-9.
- (144) Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 1994 Sep 30;265:2088-90.
- (145) Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994 Oct 7;266:66-71.
- (146) Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995 Dec 21;378:789-92.

- (147) Gayther SA, Russell P, Harrington P, et al. The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer-susceptibility genes. Am J Hum Genet 1999 Oct;65:1021-9.
- (148) Shih HA, Couch FJ, Nathanson KL, et al. BRCA1 and BRCA2 mutation frequency in women evaluated in a breast cancer risk evaluation clinic. J Clin Oncol 2002 Feb 15;20:994-9.
- (149) Oncogénétique en 2014 Consultations, laboratoires et prise en charge. Collection Appui à la décision, INCa; 2016.
- (150) Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet 2008 Jan;40:17-22.
- (151) Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 1998 Mar;62:676-89.
- (152) Cui J, Antoniou AC, Dite GS, et al. After BRCA1 and BRCA2-what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. Am J Hum Genet 2001 Feb;68:420-31.
- (153) Antoniou AC, Pharoah PD, McMullan G, et al. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol 2001 Jul;21:1-18.
- (154) Antoniou AC, Pharoah PD, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002 Jan 7;86:76-83.
- (155) Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. Oncogene 2006 Sep 25;25:5898-905.
- (156) Pharoah PD, Antoniou A, Bobrow M, et al. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002 May;31:33-6.
- (157) Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. Annu Rev Genomics Hum Genet 2008;9:321-45.
- (158) Aloraifi F, Boland MR, Green AJ, Geraghty JG. Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. Surg Oncol 2015 Jun;24:100-9.
- (159) MORTON NE. Sequential tests for the detection of linkage. Am J Hum Genet 1955 Sep;7:277-318.
- (160) MORTON NE. Logarithm of odds (lods) for linkage in complex inheritance. Proc Natl Acad Sci U S A 1996 Apr 16;93:3471-6.
- (161) Royer-Pokora B, Kunkel LM, Monaco AP, et al. Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. Nature 1986 Jul 3;322:32-8.

- (162) Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature 1998 Mar 26;392:402-5.
- (163) Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature 1998 Jan 8;391:184-7.
- (164) Hemminki A, Tomlinson I, Markie D, et al. Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. Nat Genet 1997 Jan;15:87-90.
- (165) Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet 1998 Jan;18:38-43.
- (166) Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22-23. Nat Genet 1996 May;13:114-6.
- (167) Nelen MR, van Staveren WC, Peeters EA, et al. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Hum Mol Genet 1997 Aug;6:1383-7.
- (168) Kainu T, Juo SH, Desper R, et al. Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. Proc Natl Acad Sci U S A 2000 Aug 15;97:9603-8.
- (169) Rahman N, Teare MD, Seal S, et al. Absence of evidence for a familial breast cancer susceptibility gene at chromosome 8p12-p22. Oncogene 2000 Aug 24;19:4170-3.
- (170) Thompson D, Szabo CI, Mangion J, et al. Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. Proc Natl Acad Sci U S A 2002 Jan 22;99:827-31.
- (171) Maguire P, Holmberg K, Kost-Alimova M, et al. CGH analysis of familial non-BRCA1/BRCA2 breast tumors and mutation screening of a candidate locus on chromosome 17q11.2-12. Int J Mol Med 2005 Jul;16:135-41.
- (172) Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell 2007 Feb;11:103-5.
- (173) Vahteristo P, Bartkova J, Eerola H, et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. Am J Hum Genet 2002 Aug;71:432-8.
- (174) Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxiatelangiectasia are breast cancer susceptibility alleles. Nat Genet 2006 Aug;38:873-5.
- (175) Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007 Feb;39:165-7.
- (176) Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005 Feb;6:95-108.
- (177) Strachan T, Read AP. Human molecular genetics. 2nd ed. New York: Wiley-Liss; 1999.

- (178) Dunning AM, Healey CS, Pharoah PD, et al. A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1999 Oct;8:843-54.
- (179) Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004 Nov;4:850-60.
- (180) Mardis ER. A decade's perspective on DNA sequencing technology. Nature 2011 Feb 10;470:198-203.
- (181) Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975 May 25;94:441-8.
- (182) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977 Dec;74:5463-7.
- (183) Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009 Nov 10;106:19096-101.
- (184) Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008 Oct;26:1135-45.
- (185) Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009 Sep 10;461:272-6.
- (186) Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 2012 Jan;13:36-46.
- (187) Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010 Jan;42:30-5.
- (188) Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 2013 Oct;14:681-91.
- (189) Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 2013 Oct 17;369:1502-11.
- (190) Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. PLoS Genet 2012 Sep;8:e1002894.
- (191) Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet 2012 Apr 6;90:734-9.
- (192) Cybulski C, Carrot-Zhang J, Kluzniak W, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet 2015 Jun;47:643-6.
- (193) Walsh T, Lee MK, Casadei S, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci U S A 2010 Jul 13;107:12629-33.

- (194) Castera L, Krieger S, Rousselin A, et al. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. Eur J Hum Genet 2014 Nov;22:1305-13.
- (195) Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature 2013 Jan 17;493:406-10.
- (196) Frebourg T. The challenge for the next generation of medical geneticists. Hum Mutat 2014 Aug;35:909-11.
- (197) Chen CF, Li S, Chen Y, et al. The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor. J Biol Chem 1996 Dec 20;271:32863-8.
- (198) Sy SM, Huen MS, Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci U S A 2009 Apr 28;106:7155-60.
- (199) Bork P, Hofmann K, Bucher P, et al. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J 1997 Jan;11:68-76.
- (200) Welcsh PL, Owens KN, King MC. Insights into the functions of BRCA1 and BRCA2. Trends Genet 2000 Feb;16:69-74.
- (201) Lorick KL, Jensen JP, Fang S, et al. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc Natl Acad Sci U S A 1999 Sep 28;96:11364-9.
- (202) Joukov V, Chen J, Fox EA, Green JB, Livingston DM. Functional communication between endogenous BRCA1 and its partner, BARD1, during Xenopus laevis development. Proc Natl Acad Sci U S A 2001 Oct 9;98:12078-83.
- (203) Huen MS, Sy SM, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. Nat Rev Mol Cell Biol 2010 Feb;11:138-48.
- (204) Gudmundsdottir K, Ashworth A. The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. Oncogene 2006 Sep 25;25:5864-74.
- (205) Venkitaraman AR. Functions of BRCA1 and BRCA2 in the biological response to DNA damage. J Cell Sci 2001 Oct;114:3591-8.
- (206) Silver DP, Dimitrov SD, Feunteun J, et al. Further evidence for BRCA1 communication with the inactive X chromosome. Cell 2007 Mar 9;128:991-1002.
- (207) Wong AK, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. J Biol Chem 1997 Dec 19;272:31941-4.
- (208) Carreira A, Hilario J, Amitani I, et al. The BRC repeats of BRCA2 modulate the DNA-binding selectivity of RAD51. Cell 2009 Mar 20;136:1032-43.

- (209) Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 2006 Jun 23;22:719-29.
- (210) Bryant HE, Schultz N, Thomas HD, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 2005 Apr 14;434:913-7.
- (211) Farmer H, McCabe N, Lord CJ, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005 Apr 14;434:917-21.
- (212) Fong PC, Boss DS, Yap TA, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 2009 Jul 9;361:123-34.
- (213) Easton DF, Deffenbaugh AM, Pruss D, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 2007 Nov;81:873-83.
- (214) Sharan SK, Wims M, Bradley A. Murine Brca1: sequence and significance for human missense mutations. Hum Mol Genet 1995 Dec;4:2275-8.
- (215) Thompson D, Easton D. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. Am J Hum Genet 2001 Feb;68:410-9.
- (216) Thompson D, Easton D. Variation in BRCA1 cancer risks by mutation position. Cancer Epidemiol Biomarkers Prev 2002 Apr;11:329-36.
- (217) Ware MD, DeSilva D, Sinilnikova OM, et al. Does nonsense-mediated mRNA decay explain the ovarian cancer cluster region of the BRCA2 gene? Oncogene 2006 Jan 12;25:323-8.
- (218) Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003 May;72:1117-30.
- (219) Easton DF, Matthews FE, Ford D, Swerdlow AJ, Peto J. Cancer mortality in relatives of women with ovarian cancer: the OPCS Study. Office of Population Censuses and Surveys. Int J Cancer 1996 Jan 26;65:284-94.
- (220) Ford D, Easton DF, Stratton M, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 1998 Mar;62:676-89.
- (221) Peto J, Easton DF, Matthews FE, Ford D, Swerdlow AJ. Cancer mortality in relatives of women with breast cancer: the OPCS Study. Office of Population Censuses and Surveys. Int J Cancer 1996 Jan 26;65:275-83.
- (222) Ramus SJ, Harrington PA, Pye C, et al. Contribution of BRCA1 and BRCA2 mutations to inherited ovarian cancer. Hum Mutat 2007 Dec;28:1207-15.
- (223) Thompson D, Easton DF. Cancer Incidence in BRCA1 mutation carriers. J Natl Cancer Inst 2002 Sep 18;94:1358-65.

- (224) Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst 1999 Aug 4;91:1310-6.
- (225) Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. Lancet 2006 Feb 18;367:595-604.
- (226) Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. Science 2002 Jul 26;297:606-9.
- (227) Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. J Med Genet 2007 Jan;44:1-9.
- (228) Evers B, Jonkers J. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. Oncogene 2006 Sep 25;25:5885-97.
- (229) Sawyer SL, Tian L, Kahkonen M, et al. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov 2015 Feb;5:135-42.
- (230) Domchek SM, Tang J, Stopfer J, et al. Biallelic deleterious BRCA1 mutations in a woman with early-onset ovarian cancer. Cancer Discov 2013 Apr;3:399-405.
- (231) Chompret A, Abel A, Stoppa-Lyonnet D, et al. Sensitivity and predictive value of criteria for p53 germline mutation screening. J Med Genet 2001 Jan;38:43-7.
- (232) Li FP, Fraumeni JF, Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Ann Intern Med 1969 Oct;71:747-52.
- (233) Liang L, Shao C, Deng L, et al. Radiation-induced genetic instability in vivo depends on p53 status. Mutat Res 2002 May 22;502:69-80.
- (234) Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015 Jan;43:D805-D811.
- (235) Yin Y, Tainsky MA, Bischoff FZ, Strong LC, Wahl GM. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. Cell 1992 Sep 18;70:937-48.
- (236) Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990 Nov 30;250:1233-8.
- (237) Chompret A, Brugieres L, Ronsin M, et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. Br J Cancer 2000 Jun;82:1932-7.
- (238) Gonzalez KD, Noltner KA, Buzin CH, et al. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. J Clin Oncol 2009 Mar 10;27:1250-6.
- (239) Masciari S, Dillon DA, Rath M, et al. Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. Breast Cancer Res Treat 2012 Jun;133:1125-30.
- (240) Melhem-Bertrandt A, Bojadzieva J, Ready KJ, et al. Early onset HER2-positive breast cancer is associated with germline TP53 mutations. Cancer 2012 Feb 15;118:908-13.

- (241) Wilson JR, Bateman AC, Hanson H, et al. A novel HER2-positive breast cancer phenotype arising from germline TP53 mutations. J Med Genet 2010 Nov;47:771-4.
- (242) McCuaig JM, Armel SR, Novokmet A, et al. Routine TP53 testing for breast cancer under age 30: ready for prime time? Fam Cancer 2012 Dec;11:607-13.
- (243) Mouchawar J, Korch C, Byers T, et al. Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. Cancer Res 2010 Jun 15;70:4795-800.
- (244) Tinat J, Bougeard G, Baert-Desurmont S, et al. 2009 version of the Chompret criteria for Li Fraumeni syndrome. J Clin Oncol 2009 Sep 10;27:e108-e109.
- (245) Giardiello FM, Trimbath JD. Peutz-Jeghers syndrome and management recommendations. Clin Gastroenterol Hepatol 2006 Apr;4:408-15.
- (246) Alessi DR, Sakamoto K, Bayascas JR. LKB1-dependent signaling pathways. Annu Rev Biochem 2006;75:137-63.
- (247) Hearle N, Schumacher V, Menko FH, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res 2006 May 15;12:3209-15.
- (248) van Lier MG, Wagner A, Mathus-Vliegen EM, et al. High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol 2010 Jun;105:1258-64.
- (249) Maehama T, Dixon JE. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem 1998 May 29;273:13375-8.
- (250) Cantley LC, Neel BG. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. Proc Natl Acad Sci U S A 1999 Apr 13;96:4240-5.
- (251) Nelen MR, Kremer H, Konings IB, et al. Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. Eur J Hum Genet 1999 Apr;7:267-73.
- (252) Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22-23. Nat Genet 1996 May;13:114-6.
- (253) Hsu SC, Volpert OV, Steck PA, et al. Inhibition of angiogenesis in human glioblastomas by chromosome 10 induction of thrombospondin-1. Cancer Res 1996 Dec 15;56:5684-91.
- (254) Li J, Yen C, Liaw D, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997 Mar 28;275:1943-7.
- (255) Steck PA, Pershouse MA, Jasser SA, et al. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 1997 Apr;15:356-62.

- (256) Li DM, Sun H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. Cancer Res 1997 Jun 1;57:2124-9.
- (257) Liaw D, Marsh DJ, Li J, et al. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 1997 May;16:64-7.
- (258) Blumenthal GM, Dennis PA. PTEN hamartoma tumor syndromes. Eur J Hum Genet 2008 Nov;16:1289-300.
- (259) Bubien V, Bonnet F, Brouste V, et al. High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. J Med Genet 2013 Apr;50:255-63.
- (260) Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature 1998 Mar 26;392:402-5.
- (261) Guilford PJ, Hopkins JB, Grady WM, et al. E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. Hum Mutat 1999;14:249-55.
- (262) Overduin M, Harvey TS, Bagby S, et al. Solution structure of the epithelial cadherin domain responsible for selective cell adhesion. Science 1995 Jan 20;267:386-9.
- (263) Keller G, Vogelsang H, Becker I, et al. Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. Am J Pathol 1999 Aug;155:337-42.
- (264) Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 2001 Dec;121:1348-53.
- (265) Hansford S, Kaurah P, Li-Chang H, et al. Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. JAMA Oncol 2015 Apr;1:23-32.
- (266) Masciari S, Larsson N, Senz J, et al. Germline E-cadherin mutations in familial lobular breast cancer. J Med Genet 2007 Nov;44:726-31.
- (267) Lalloo F, Varley J, Moran A, et al. BRCA1, BRCA2 and TP53 mutations in very early-onset breast cancer with associated risks to relatives. Eur J Cancer 2006 May;42:1143-50.
- (268) Evans DG, Birch JM, Thorneycroft M, et al. Low rate of TP53 germline mutations in breast cancer/sarcoma families not fulfilling classical criteria for Li-Fraumeni syndrome. J Med Genet 2002 Dec;39:941-4.
- (269) Borresen AL, Andersen TI, Garber J, et al. Screening for germ line TP53 mutations in breast cancer patients. Cancer Res 1992 Jun 1;52:3234-6.
- (270) Bignell GR, Barfoot R, Seal S, et al. Low frequency of somatic mutations in the LKB1/Peutz-Jeghers syndrome gene in sporadic breast cancer. Cancer Res 1998 Apr 1;58:1384-6.

- (271) Chen J, Lindblom P, Lindblom A. A study of the PTEN/MMAC1 gene in 136 breast cancer families. Hum Genet 1998 Jan;102:124-5.
- (272) Guenard F, Labrie Y, Ouellette G, et al. Germline mutations in the breast cancer susceptibility gene PTEN are rare in high-risk non-BRCA1/2 French Canadian breast cancer families. Fam Cancer 2007;6:483-90.
- (273) Shugart YY, Cour C, Renard H, et al. Linkage analysis of 56 multiplex families excludes the Cowden disease gene PTEN as a major contributor to familial breast cancer. J Med Genet 1999 Sep;36:720-1.
- (274) Gatti RA, Berkel I, Boder E, et al. Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. Nature 1988 Dec 8;336:577-80.
- (275) Savitsky K, Bar-Shira A, Gilad S, et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 1995 Jun 23;268:1749-53.
- (276) Abraham RT. PI 3-kinase related kinases: 'big' players in stress-induced signaling pathways. DNA Repair (Amst) 2004 Aug;3:883-7.
- (277) Shiloh Y. ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer 2003 Mar;3:155-68.
- (278) Ahmed M, Rahman N. ATM and breast cancer susceptibility. Oncogene 2006 Sep 25;25:5906-11.
- (279) Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet 2001 Mar;27:247-54.
- (280) Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nat Rev Mol Cell Biol 2013 Apr;14:197-210.
- (281) Shiloh Y. ATM: expanding roles as a chief guardian of genome stability. Exp Cell Res 2014 Nov 15;329:154-61.
- (282) Lavin MF, Kozlov S. ATM activation and DNA damage response. Cell Cycle 2007 Apr 15;6:931-42.
- (283) Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature 2003 Jan 30;421:499-506.
- (284) Choi M, Kipps T, Kurzrock R. ATM Mutations in Cancer: Therapeutic Implications. Mol Cancer Ther 2016 Aug;15:1781-91.
- (285) Matsuoka S, Ballif BA, Smogorzewska A, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 2007 May 25;316:1160-6.
- (286) Ditch S, Paull TT. The ATM protein kinase and cellular redox signaling: beyond the DNA damage response. Trends Biochem Sci 2012 Jan;37:15-22.

- (287) Mu JJ, Wang Y, Luo H, et al. A proteomic analysis of ataxia telangiectasia-mutated (ATM)/ATM-Rad3-related (ATR) substrates identifies the ubiquitin-proteasome system as a regulator for DNA damage checkpoints. J Biol Chem 2007 Jun 15;282:17330-4.
- (288) Bensimon A, Schmidt A, Ziv Y, et al. ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. Sci Signal 2010;3:rs3.
- (289) Abraham RT. Cell cycle checkpoint signaling through the ATM and ATR kinases. Genes Dev 2001 Sep 1;15:2177-96.
- (290) Barlow C, Dennery PA, Shigenaga MK, et al. Loss of the ataxia-telangiectasia gene product causes oxidative damage in target organs. Proc Natl Acad Sci U S A 1999 Aug 17;96:9915-9.
- (291) Chun HH, Gatti RA. Ataxia-telangiectasia, an evolving phenotype. DNA Repair (Amst) 2004 Aug;3:1187-96.
- (292) Taylor AM, Groom A, Byrd PJ. Ataxia-telangiectasia-like disorder (ATLD)-its clinical presentation and molecular basis. DNA Repair (Amst) 2004 Aug;3:1219-25.
- (293) Taylor AM, Metcalfe JA, Thick J, Mak YF. Leukemia and lymphoma in ataxia telangiectasia. Blood 1996 Jan 15;87:423-38.
- (294) Li A, Swift M. Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients. Am J Med Genet 2000 May 29;92:170-7.
- (295) Swift M, Morrell D, Cromartie E, et al. The incidence and gene frequency of ataxiatelangiectasia in the United States. Am J Hum Genet 1986 Nov;39:573-83.
- (296) Easton DF. Cancer risks in A-T heterozygotes. Int J Radiat Biol 1994 Dec;66:S177-S182.
- (297) Swift M, Reitnauer PJ, Morrell D, Chase CL. Breast and other cancers in families with ataxia-telangiectasia. N Engl J Med 1987 May 21;316:1289-94.
- (298) Swift M, Sholman L, Perry M, Chase C. Malignant neoplasms in the families of patients with ataxia-telangiectasia. Cancer Res 1976 Jan;36:209-15.
- (299) Athma P, Rappaport R, Swift M. Molecular genotyping shows that ataxiatelangiectasia heterozygotes are predisposed to breast cancer. Cancer Genet Cytogenet 1996 Dec;92:130-4.
- (300) Janin N, Andrieu N, Ossian K, et al. Breast cancer risk in ataxia telangiectasia (AT) heterozygotes: haplotype study in French AT families. Br J Cancer 1999 Jun;80:1042-5.
- (301) Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. N Engl J Med 1991 Dec 26;325:1831-6.
- (302) Thompson D, Duedal S, Kirner J, et al. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst 2005 Jun 1;97:813-22.

- (303) Bebb DG, Yu Z, Chen J, et al. Absence of mutations in the ATM gene in forty-seven cases of sporadic breast cancer. Br J Cancer 1999 Aug;80:1979-81.
- (304) Bishop DT, Hopper J. AT-tributable risks? Nat Genet 1997 Mar;15:226.
- (305) Chen J, Birkholtz GG, Lindblom P, Rubio C, Lindblom A. The role of ataxiatelangiectasia heterozygotes in familial breast cancer. Cancer Res 1998 Apr 1;58:1376-9.
- (306) FitzGerald MG, Bean JM, Hegde SR, et al. Heterozygous ATM mutations do not contribute to early onset of breast cancer. Nat Genet 1997 Mar;15:307-10.
- (307) Izatt L, Greenman J, Hodgson S, et al. Identification of germline missense mutations and rare allelic variants in the ATM gene in early-onset breast cancer. Genes Chromosomes Cancer 1999 Dec;26:286-94.
- (308) Olsen JH, Hahnemann JM, Borresen-Dale AL, et al. Breast and other cancers in 1445 blood relatives of 75 Nordic patients with ataxia telangiectasia. Br J Cancer 2005 Jul 25;93:260-5.
- (309) Vorechovsky I, Rasio D, Luo L, et al. The ATM gene and susceptibility to breast cancer: analysis of 38 breast tumors reveals no evidence for mutation. Cancer Res 1996 Jun 15;56:2726-32.
- (310) Angele S, Hall J. The ATM gene and breast cancer: is it really a risk factor? Mutat Res 2000 Apr;462:167-78.
- (311) Khanna KK, Chenevix-Trench G. ATM and genome maintenance: defining its role in breast cancer susceptibility. J Mammary Gland Biol Neoplasia 2004 Jul;9:247-62.
- (312) Hall J. The Ataxia-telangiectasia mutated gene and breast cancer: gene expression profiles and sequence variants. Cancer Lett 2005 Sep 28;227:105-14.
- (313) Gilardini Montani MS, Prodosmo A, Stagni V, et al. ATM-depletion in breast cancer cells confers sensitivity to PARP inhibition. J Exp Clin Cancer Res 2013;32:95.
- (314) Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell 2003 May;3:421-9.
- (315) Bell DW, Varley JM, Szydlo TE, et al. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science 1999 Dec 24;286:2528-31.
- (316) Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 2002 May;31:55-9.
- (317) CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. Am J Hum Genet 2004 Jun;74:1175-82.
- (318) Wang Y, Dai B, Ye D. CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis. Int J Clin Exp Med 2015;8:15708-15.

- (319) Cybulski C, Gorski B, Huzarski T, et al. CHEK2 is a multiorgan cancer susceptibility gene. Am J Hum Genet 2004 Dec;75:1131-5.
- (320) Dong X, Wang L, Taniguchi K, et al. Mutations in CHEK2 associated with prostate cancer risk. Am J Hum Genet 2003 Feb;72:270-80.
- (321) Thompson D, Seal S, Schutte M, et al. A multicenter study of cancer incidence in CHEK2 1100delC mutation carriers. Cancer Epidemiol Biomarkers Prev 2006 Dec;15:2542-5.
- (322) Kilpivaara O, Vahteristo P, Falck J, et al. CHEK2 variant I157T may be associated with increased breast cancer risk. Int J Cancer 2004 Sep 10;111:543-7.
- (323) Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet 2012 Jun 15;21:2738-44.
- (324) Bogdanova N, Enssen-Dubrowinskaja N, Feshchenko S, et al. Association of two mutations in the CHEK2 gene with breast cancer. Int J Cancer 2005 Aug 20;116:263-6.
- (325) Schutte M, Seal S, Barfoot R, et al. Variants in CHEK2 other than 1100delC do not make a major contribution to breast cancer susceptibility. Am J Hum Genet 2003 Apr;72:1023-8.
- (326) Dufault MR, Betz B, Wappenschmidt B, et al. Limited relevance of the CHEK2 gene in hereditary breast cancer. Int J Cancer 2004 Jun 20;110:320-5.
- (327) Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA 2006 Mar 22;295:1379-88.
- (328) Cantor SB, Bell DW, Ganesan S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell 2001 Apr 6;105:149-60.
- (329) Peng M, Litman R, Jin Z, Fong G, Cantor SB. BACH1 is a DNA repair protein supporting BRCA1 damage response. Oncogene 2006 Apr 6;25:2245-53.
- (330) Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet 2006 Nov;38:1239-41.
- (331) Lewis AG, Flanagan J, Marsh A, et al. Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer. Breast Cancer Res 2005;7:R1005-R1016.
- (332) Frank B, Hemminki K, Meindl A, et al. BRIP1 (BACH1) variants and familial breast cancer risk: a case-control study. BMC Cancer 2007;7:83.
- (333) Guenard F, Labrie Y, Ouellette G, et al. Mutational analysis of the breast cancer susceptibility gene BRIP1 /BACH1/FANCJ in high-risk non-BRCA1/BRCA2 breast cancer families. J Hum Genet 2008;53:579-91.

- (334) Karppinen SM, Vuosku J, Heikkinen K, Allinen M, Winqvist R. No evidence of involvement of germline BACH1 mutations in Finnish breast and ovarian cancer families. Eur J Cancer 2003 Feb;39:366-71.
- (335) Kuusisto KM, Bebel A, Vihinen M, Schleutker J, Sallinen SL. Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. Breast Cancer Res 2011;13:R20.
- (336) Luo L, Lei H, Du Q, et al. No mutations in the BACH1 gene in BRCA1 and BRCA2 negative breast-cancer families linked to 17q22. Int J Cancer 2002 Apr 1;98:638-9.
- (337) Rutter JL, Smith AM, Davila MR, et al. Mutational analysis of the BRCA1-interacting genes ZNF350/ZBRK1 and BRIP1/BACH1 among BRCA1 and BRCA2-negative probands from breast-ovarian cancer families and among early-onset breast cancer cases and reference individuals. Hum Mutat 2003 Aug;22:121-8.
- (338) Vahteristo P, Yliannala K, Tamminen A, et al. BACH1 Ser919Pro variant and breast cancer risk. BMC Cancer 2006;6:19.
- (339) Cao AY, Huang J, Hu Z, et al. Mutation analysis of BRIP1/BACH1 in BRCA1/BRCA2 negative Chinese women with early onset breast cancer or affected relatives. Breast Cancer Res Treat 2009 May;115:51-5.
- (340) Catucci I, Milgrom R, Kushnir A, et al. Germline mutations in BRIP1 and PALB2 in Jewish high cancer risk families. Fam Cancer 2012 Sep;11:483-91.
- (341) Levitus M, Waisfisz Q, Godthelp BC, et al. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat Genet 2005 Sep;37:934-5.
- (342) Levran O, Attwooll C, Henry RT, et al. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. Nat Genet 2005 Sep;37:931-3.
- (343) Litman R, Peng M, Jin Z, et al. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. Cancer Cell 2005 Sep;8:255-65.
- (344) Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 2006 Jun 23;22:719-29.
- (345) Xia B, Dorsman JC, Ameziane N, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet 2007 Feb;39:159-61.
- (346) Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet 2007 Feb;39:162-4.
- (347) Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med 2014 Aug 7;371:497-506.
- (348) Erkko H, Xia B, Nikkila J, et al. A recurrent mutation in PALB2 in Finnish cancer families. Nature 2007 Mar 15;446:316-9.

- (349) Foulkes WD, Ghadirian P, Akbari MR, et al. Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. Breast Cancer Res 2007;9:R83.
- (350) Zhu XD, Kuster B, Mann M, Petrini JH, de LT. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. Nat Genet 2000 Jul;25:347-52.
- (351) Stracker TH, Theunissen JW, Morales M, Petrini JH. The Mre11 complex and the metabolism of chromosome breaks: the importance of communicating and holding things together. DNA Repair (Amst) 2004 Aug;3:845-54.
- (352) Heikkinen K, Rapakko K, Karppinen SM, et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis 2006 Aug;27:1593-9.
- (353) Tommiska J, Seal S, Renwick A, et al. Evaluation of RAD50 in familial breast cancer predisposition. Int J Cancer 2006 Jun 1;118:2911-6.
- (354) Bogdanova N, Feshchenko S, Schurmann P, et al. Nijmegen Breakage Syndrome mutations and risk of breast cancer. Int J Cancer 2008 Feb 15;122:802-6.
- (355) Bartkova J, Tommiska J, Oplustilova L, et al. Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. Mol Oncol 2008 Dec;2:296-316.
- (356) Stewart GS, Maser RS, Stankovic T, et al. The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell 1999 Dec 10;99:577-87.
- (357) Damiola F, Pertesi M, Oliver J, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutationscreening study. Breast Cancer Res 2014;16:R58.
- (358) Thompson ER, Boyle SE, Johnson J, et al. Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. Hum Mutat 2012 Jan;33:95-9.
- (359) Pelttari LM, Heikkinen T, Thompson D, et al. RAD51C is a susceptibility gene for ovarian cancer. Hum Mol Genet 2011 Aug 15;20:3278-88.
- (360) Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet 2010 May;42:410-4.
- (361) Blanco A, Gutierrez-Enriquez S, Santamarina M, et al. RAD51C germline mutations found in Spanish site-specific breast cancer and breast-ovarian cancer families. Breast Cancer Res Treat 2014 Aug;147:133-43.

- (362) Rashid MU, Muhammad N, Faisal S, Amin A, Hamann U. Deleterious RAD51C germline mutations rarely predispose to breast and ovarian cancer in Pakistan. Breast Cancer Res Treat 2014 Jun;145:775-84.
- (363) Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996 Sep 13;273:1516-7.
- (364) Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007 Jun 28;447:1087-93.
- (365) Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007 Jul;39:870-4.
- (366) Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007 Jul;39:865-9.
- (367) Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007 Mar;39:352-8.
- (368) Antoniou AC, Spurdle AB, Sinilnikova OM, et al. Common breast cancerpredisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. Am J Hum Genet 2008 Apr;82:937-48.
- (369) Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol 2011 May;12:477-88.
- (370) Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet 2015 Apr;47:373-80.
- (371) Varghese JS, Easton DF. Genome-wide association studies in common cancers--what have we learnt? Curr Opin Genet Dev 2010 Jun;20:201-9.
- (372) McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010 Apr 16;141:210-7.
- (373) Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common diseaseassociated variation in regulatory DNA. Science 2012 Sep 7;337:1190-5.
- (374) Pomerantz MM, Ahmadiyeh N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 2009 Aug;41:882-4.
- (375) Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Mol Cell Biol 2010 Mar;30:1411-20.
- (376) Fanale D, Amodeo V, Corsini LR, et al. Breast cancer genome-wide association studies: there is strength in numbers. Oncogene 2012 Apr 26;31:2121-8.

- (377) Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. Ann Oncol 2011 Jan;22 Suppl 1:i11-i17.
- (378) Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res 2006 Sep;16:1136-48.
- (379) Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
- (380) Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 2007 Mar 15;23:657-63.
- (381) Letouze E, Sow A, Petel F, et al. Identity by descent mapping of founder mutations in cancer using high-resolution tumor SNP data. PLoS One 2012;7:e35897.
- (382) Gusev A, Lowe JK, Stoffel M, et al. Whole population, genome-wide mapping of hidden relatedness. Genome Res 2009 Feb;19:318-26.
- (383) Gusev A, Kenny EE, Lowe JK, et al. DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. Am J Hum Genet 2011 Jun 10;88:706-17.
- (384) Priya RR, Rajasimha HK, Brooks MJ, Swaroop A. Exome sequencing: capture and sequencing of all human coding regions for disease gene discovery. Methods Mol Biol 2012;884:335-51.
- (385) Bainbridge MN, Wang M, Wu Y, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol 2011;12:R68.
- (386) Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009 Feb;27:182-9.
- (387) Chang CT, Tsai CN, Tang CY, et al. Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. ScientificWorldJournal 2012;2012:365104.
- (388) Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010 Mar 1;26:589-95.
- (389) McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010 Sep;20:1297-303.
- (390) McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122.
- (391) Underwood SL, Christoforou A, Thomson PA, et al. Association analysis of the chromosome 4p-located G protein-coupled receptor 78 (GPR78) gene in bipolar affective disorder and schizophrenia. Mol Psychiatry 2006 Apr;11:384-94.

- (392) Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010 Apr;7:248-9.
- (393) Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.
- (394) Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010 Aug;7:575-6.
- (395) Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006 Apr;43:295-305.
- (396) Kawarai T, Tajima A, Kuroda Y, et al. A homozygous mutation of VWA3B causes cerebellar ataxia with intellectual disability. J Neurol Neurosurg Psychiatry 2016 Jun;87:656-62.
- (397) Dittrich N, Berrocal-Almanza LC, Thada S, et al. Toll-like receptor 1 variations influence susceptibility and immune response to Mycobacterium tuberculosis. Tuberculosis (Edinb) 2015 May;95:328-35.
- (398) Sherman J, Tester DJ, Ackerman MJ. Targeted mutational analysis of ankyrin-B in 541 consecutive, unrelated patients referred for long QT syndrome genetic testing and 200 healthy subjects. Heart Rhythm 2005 Nov;2:1218-23.
- (399) Wu J, Lei G, Mei M, Tang Y, Li H. A novel C53/LZAP-interacting protein regulates stability of C53/LZAP and DDRGK domain-containing Protein 1 (DDRGK1) and modulates NF-kappaB signaling. J Biol Chem 2010 May 14;285:15126-36.
- (400) Dagoneau N, Goulet M, Genevieve D, et al. DYNC2H1 mutations cause asphyxiating thoracic dystrophy and short rib-polydactyly syndrome, type III. Am J Hum Genet 2009 May;84:706-11.
- (401) Nioi P, Sigurdsson A, Thorleifsson G, et al. Variant ASGR1 Associated with a Reduced Risk of Coronary Artery Disease. N Engl J Med 2016 Jun 2;374:2131-41.
- (402) Seino Y, Miki T, Kiyonari H, et al. Isx participates in the maintenance of vitamin A metabolism by regulation of beta-carotene 15,15'-monooxygenase (Bcmo1) expression. J Biol Chem 2008 Feb 22;283:4905-11.
- (403) Yang H, Sasaki T, Minoshima S, Shimizu N. Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. Genomics 2007 Aug;90:249-60.
- (404) Lee IK, Kim KS, Kim H, et al. MAP, a protein interacting with a tumor suppressor, merlin, through the run domain. Biochem Biophys Res Commun 2004 Dec 17;325:774-83.
- (405) Allinen M, Launonen V, Laake K, et al. ATM mutations in Finnish breast cancer patients. J Med Genet 2002 Mar;39:192-6.

- (406) Pylkas K, Tommiska J, Syrjakoski K, et al. Evaluation of the role of Finnish ataxiatelangiectasia mutations in hereditary predisposition to breast cancer. Carcinogenesis 2007 May;28:1040-5.
- (407) Vorechovsky I, Luo L, Lindblom A, et al. ATM mutations in cancer families. Cancer Res 1996 Sep 15;56:4130-3.
- (408) Easton DF, Pharoah PD, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med 2015 Jun 4;372:2243-57.
- (409) Finishing the euchromatic sequence of the human genome. Nature 2004 Oct 21;431:931-45.
- (410) Vissers LE, de LJ, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet 2010 Dec;42:1109-12.
- (411) Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010 Sep;42:790-3.
- (412) Dorschner MO, Amendola LM, Turner EH, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet 2013 Oct 3;93:631-40.
- (413) Nakashima M, Takano K, Osaka H, et al. Causative novel PNKP mutations and concomitant PCDH15 mutations in a patient with microcephaly with early-onset seizures and developmental delay syndrome and hearing loss. J Hum Genet 2014 Aug;59:471-4.
- (414) O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 2011 Jun;43:585-9.
- (415) Hoischen A, van Bon BW, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet 2010 Jun;42:483-5.
- (416) Wang JL, Yang X, Xia K, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 2010 Dec;133:3510-8.
- (417) Lai CC, Yeh YH, Hsieh WP, et al. Whole-exome sequencing to identify a novel LMNA gene mutation associated with inherited cardiac conduction disease. PLoS One 2013;8:e83322.
- (418) Harms MB, Sommerville RB, Allred P, et al. Exome sequencing reveals DNAJB6 mutations in dominantly-inherited myopathy. Ann Neurol 2012 Mar;71:407-16.
- (419) Aloraifi F, McDevitt T, Martiniano R, et al. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. FEBS J 2015 Sep;282:3424-37.
- (420) Byers H, Wallis Y, van Veen EM, et al. Sensitivity of BRCA1/2 testing in high-risk breast/ovarian/male breast cancer families: little contribution of comprehensive RNA/NGS panel testing. Eur J Hum Genet 2016 Jun 8.
- (421) Ellingson MS, Hart SN, Kalari KR, et al. Exome sequencing reveals frequent deleterious germline variants in cancer susceptibility genes in women with invasive

breast cancer undergoing neoadjuvant chemotherapy. Breast Cancer Res Treat 2015 Sep;153:435-43.

- (422) Maxwell KN, Wubbenhorst B, D'Andrea K, et al. Prevalence of mutations in a panel of breast cancer susceptibility genes in BRCA1/2-negative patients with early-onset breast cancer. Genet Med 2015 Aug;17:630-8.
- (423) Snape K, Ruark E, Tarpey P, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res Treat 2012 Jul;134:429-33.
- (424) Maxwell KN, Hart SN, Vijai J, et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet 2016 May 5;98:801-17.
- (425) Broeks A, Urbanus JH, Floore AN, et al. ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. Am J Hum Genet 2000 Feb;66:494-500.
- (426) Roberts NJ, Jiao Y, Yu J, et al. ATM mutations in hereditary pancreatic cancer patients. Cancer Discov 2012 Jan;2:41-6.
- (427) Rio PG, Pernin D, Bay JO, et al. Loss of heterozygosity of BRCA1, BRCA2 and ATM genes in sporadic invasive ductal breast carcinoma. Int J Oncol 1998 Oct;13:849-53.
- (428) Bernstein JL, Haile RW, Stovall M, et al. Radiation exposure, the ATM Gene, and contralateral breast cancer in the women's environmental cancer and radiation epidemiology study. J Natl Cancer Inst 2010 Apr 7;102:475-83.
- (429) Renaux-Petel M, Sesboue R, Baert-Desurmont S, et al. The MDM2 285G-309G haplotype is associated with an earlier age of tumour onset in patients with Li-Fraumeni syndrome. Fam Cancer 2014 Mar;13:127-30.
- (430) Morak M, Massdorf T, Sykora H, Kerscher M, Holinski-Feder E. First evidence for digenic inheritance in hereditary colorectal cancer by mutations in the base excision repair genes. Eur J Cancer 2011 May;47:1046-55.
- (431) Yoo HM, Kang SH, Kim JY, et al. Modification of ASC1 by UFM1 is crucial for ERalpha transactivation and breast cancer development. Mol Cell 2014 Oct 23;56:261-74.

7 URL

Align GVGD: <u>http://agvgd.hci.utah.edu/agvgd_input.php</u>

BRCA Share[™] (formerly UMD-BRCA1 mutations database): <u>http://www.umd.be/BRCA1/</u> BRCA Share[™] (formerly UMD-BRCA2 mutations database): <u>http://www.umd.be/BRCA2/</u> Catalogue of Somatic Mutations In Cancer (COSMIC): <u>http://cancer.sanger.ac.uk/cosmic/</u> Exome Variant Server (EVS), NHLBI Exome Sequencing Project (ESP), Seattle, WA: <u>http://evs.gs.washington.edu/EVS/</u>

Exome Aggregation Consortium (ExAC), Cambridge, MA: http://exac.broadinstitute.org

MutationTaster: <u>http://www.mutationtaster.org/</u> OMIM® (Online Mendelian Inheritance in Man): <u>http://www.omim.org/</u> PolyPhen-2: <u>http://genetics.bwh.harvard.edu/pph2/</u> SIFT: <u>http://sift.bii.a-star.edu.sg/www/SIFT_aligned_seqs_submit.html</u>

8 ANNEXES




Annexe 2: Arbre généalogique de la famille G1996019.

que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros «HS» indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.





Annexe 3: Arbre généalogique de la famille G2001185.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros «HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.



Annexe 4: Arbre généalogique de la famille G2003221 (Témoin BRCAI).

individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour 'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.

A noter dans cette famille, l'identification d'une mutation constitutionnelle délétère du gène BRCA1 chez le cas index (patiente II:17).





Annexe 5: Arbre généalogique de la famille G2007180.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros «HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) Signification des symboles : symbole barré, individu décédé; symbole entre crochets, individu adopté; symbole entièrement grisé, pour l'individu correspondant. Les cas index sont indiqués par une flèche.

individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.

218





Annexe 6: Arbre généalogique de la famille G2009223.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.





Annexe 7: Arbre généalogique de la famille G1995064.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

présenté un autre cancer que le cancer du sein.



Annexe 8: Arbre généalogique de la famille G2007398.

« HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entre crochets, individu adopté; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.



Annexe 9: Arbre généalogique de la famille G2010220.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entre crochets, individu adopté; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.



Annexe 10: Arbre généalogique de la famille G2009373 (Témoin BRCA2).

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.

A noter l'identification dans cette famille, d'une mutation constitutionnelle délétère du gène BRCA2.



Annexe 11: Arbre généalogique de la famille G1995037.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein; symbole avec pointillés gris, individu ayant présenté une tumeur bénigne.



Annexe 12: Arbre généalogique de la famille G1995042.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entre crochets, individu adopté; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.





Annexe 13: Arbre généalogique de la famille G1995079.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche. individu ayant présenté un autre cancer que le cancer du sein.



Annexe 14: Arbre généalogique de la famille G2002199.

de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein; symbole avec pointillés gris, individu ayant présenté une tumeur bénigne.



Annexe 15: Arbre généalogique de la famille G2006232.

228

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

grisé, individu ayant présenté un autre cancer que le cancer du sein.



Annexe 16: Arbre généalogique de la famille G2009159.

même que le type de cancer et l'âge au diagnostic lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.



Annexe 17: Arbre généalogique de la famille G2010325.

L'âge indiqué sous chaque individu est l'âge aux dernières nouvelles ou l'âge du décès. La survenue d'un cancer est indiquée sous chaque individu, de même que le type de cancer et l'âge au diagnostic, lorsque ces informations étaient disponibles. Les numéros « HS » indiqués à côté de ou sous certains individus représentent la disponibilité d'un échantillon tumoral (T) et/ou d'un échantillon sanguin (N) pour l'individu correspondant. Les cas index sont indiqués par une flèche.

Signification des symboles : symbole barré, individu décédé; symbole entièrement grisé, individu ayant présenté un cancer du sein; symbole à moitié grisé, individu ayant présenté un autre cancer que le cancer du sein.

Exon	Nom de l'amorce	Séquence amorce (5'-> 3')	Position génomique*	Longeur (Nt)	Tm (°C)	% GC	Température d'hybridation (°C)	Concentration MgCl ₂ (mM)	Taille amplicon (pb)
	ATM_ex10.01F	GTCTATATATTTCCTTTTAGTTTG	108121356-108121379	24	60	25			
10	ATM_ex10.01R	CCTCTTGTCTTGACACAATGC	108121579-108121599	21	62	48	58	1,5	244
13	ATM_ex13.02F	ACTATTTCTTCAGACAACTTTTG	108124592-108124614	23	60	30	59	1 5	250
	ATM_ex13.02R	CCCTATTTCTCCTTCCTAAC	108124822-108124841	20	58	45	58	1,5	250
15	ATM_ex15F	GTATGACTACGTGGAACTTCT	108128050-108128070	21	60	42	50	1 5	500
	ATM_ex15R	CCAGTGATCAATTCCCTGCA	108128539-108128558	20	60	50	58	1,5	509
26	ATM_ex26.01F	TAATGCTGATGGTATTAAAACAG	108154855-108154877	23	60	30	50	1 5	202
	ATM_ex26.01R	CTCTCTTTGCTGTGCCATCC	108155127-108155146	20	62	55	58	1,5	292
31	ATM_ex31F	ACAGTAAGTTTTGTTGGCTTAC	108163968-108163989	22	60	36	50	1 5	214
	ATM_ex31R	GCTAGAGCATTACAGATTTTTG	108164260-108164281	22	60	36	58	1,5	314
51	ATM_ex51F	GCTAATAGAGGAGCACTGTC	108202071-108202090	20	60	46	60	1 Г	210
	ATM_ex51R	TTTGTATTTCCATTTCTTAGAGG	108202367-108202389	23	60	29	60	1,5	213
59	ATM_ex59.01F	TATGCACATCATTTAAGTAGGC	108217865-108217886	22	60	36	F 9	1 Г	216
	ATM_ex59.01R	GTACAAGTTCTGCTGACTGC	108218061-108218080	20	60	50	58	1,5	210
62	ATM_ex62F	TACTGGTTCTACTGTTTCTAAG	108235734-108235755	22	60	36		1 Г	202
	ATM_ex62R	AGTGAGGTGAACAGTTTAAAGG	108236005-108236026	22	62	40	20	1,5	293

Annexe 18. Caractéristiques des couples d'amorces utilisés pour l'amplification des exons du gène ATM

* selon transcrit hg19 GRCh37 NM_007294.3; Nt, Nucléotide; °C, degré Celsius; % GC, pourcentage de GC contenu dans l'amorce; mM, milliMolaire; pb, paire de bases.

Réactifs	Quantité* / tube		
Tampon (1X)	2,5µl		
MgCl2 (25mM)/MgSO4	1,5µl		
Amorce 5' (12 <i>,</i> 5μM)	1µl		
Amorce 3' (12,5μM)	1µl		
dNTP	0,5µl		
Taq Gold/Optimase*	0,2µl		
H2O	15,3µl		
ADN (20ng/µl)	3µl		

Annexe 19. Composition du mix réactionnel des PCR

* Pour un volume réactionnel final = 25µl

Etapes PCR *	Taq Gold	
Activation	95°C - 10mn	
Dénaturation	95°C - 30 sec	1
Hybridation**	58 ou 60°C - 30 sec	x 40 cycles
Elongation	72°C - 45 sec	I
Elongation finale	72°C - 20 mn	

Annexe 20. Programme des PCR du gène ATM

* Appareil utilisé : SimpliAmp (Life Technologies); ** Température d'hybridation dépendante de l'exon amplifié : 60°C pour l'exon 51 et 58°C pour tous les autres.

	Annexe 21.	Composition	du mix des	réactions	de séquence
--	------------	-------------	------------	-----------	-------------

Réactifs	Quantité* / tube
Sequencing Buffer	3,5µl
BigDye Terminator	0,5µl
Amorce 5' ou 3'	0,5µl
H2O	11,5µl
PCR Séquence (dilution 1/10)	4µl

* Pour un volume réactionnel final = 20µl