

**Université Bordeaux 2 – Victor Segalen**

**Ecole doctorale Sciences de la Vie**

Année 2013

N°2122

**Thèse pour le**

**DOCTORAT DE L'UNIVERSITE BORDEAUX 2**

**Mention : Sciences, Technologie et Santé**

**Option : Génétique**

Présentée et soutenue publiquement le 18 décembre 2013

Par

Audrey ROUAULT GROS

Née le 7 Septembre 1985 à Lagny-Sur-Marne

\*\*\*\*\*

ETUDE GENOMIQUE DES CANCERS DU SEIN FAMILIAUX LIES A UNE MUTATION  
CONSTITUTIONNELLE DU GENE *BRCA2*

\*\*\*\*\*

Thèse dirigée par le Docteur Nicolas SEVENET

**Composition du jury**

Mr le Professeur François MOREAU-GAUDRY,	Président
Mr le Docteur Daniel BIRNBAUM,	Rapporteur
Mr le Professeur Stéphane BEZIEAU,	Rapporteur
Mr le Professeur Pierre DUBUS,	Examineur
Mr le Docteur Marc DEBLED,	Examineur
Mr le Docteur Nicolas SEVENET,	Directeur de thèse

# REMERCIEMENTS

---

Les travaux présentés dans cette thèse ont été réalisés au sein du laboratoire INSERM U916 «Validation et Identification de Nouvelles cibles en Oncologie» dirigé par le Professeur Josy Reiffers que je remercie pour m'avoir accueillie dans son laboratoire. J'adresse également mes remerciements au CIT de la Ligue contre le Cancer.

Je remercie mon directeur de thèse, le Docteur Nicolas Sévenet pour son soutien, sa bonne humeur et son optimisme à toute épreuve.

Pour l'honneur qu'ils me font en faisant partie de ce jury, je remercie profondément le Professeur François Moreau-Gaudry qui a accepté de le présider, le Docteur Daniel Birnbaum et le Professeur Stéphane Béziau qui ont bien voulu juger ce travail, et le Professeur Pierre Dubus et le Docteur Marc Debled qui ont accepté de l'examiner.

Je remercie également le Docteur Michel Longy, le Professeur Richard Iggo, le Docteur Gaëtan MacGrogan et Carlo Lucchesi pour leur aide précieuse et leur contribution.

Il me tenait également à cœur de remercier le Professeur Benoît Arveiler, le Professeur Jean-Philippe Merlio et le Professeur Robert Saura pour la confiance et la bienveillance qu'ils m'ont accordé.

Je tiens à remercier l'ensemble de l'équipe du Département de Pathologie et du Centre de Ressources Biologiques de l'Institut Bergonié, pour leur accueil et leur enseignement. Merci Françoise, Virginie et Emmanuelle pour votre aide et nos petites discussions ! Justine et Céline, vous savez combien je vous suis reconnaissante pour le SAV, merci !

Merci Natalie, Mélyssa, Gaëlle, Eglantine et Elodie pour votre soutien !!

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

Je remercie mes amis de Paris, de Bordeaux et d'ailleurs pour tous ces bons moments partagés et les futurs.

Mes plus grands remerciements sont pour toi Kevin, merci pour ta patience et ton aide.

A tous ceux qui n'ont pas été nominalement mentionnés dans ces pages, mais qui ont contribué à la réalisation de cette thèse directement ou indirectement, je les remercie.

# RESUME

---

L'altération constitutionnelle des gènes *BRCA1* et *BRCA2* est détectée dans 20 à 30% des formes familiales de cancer du sein. La fréquence de mise en évidence d'une mutation *BRCA2* selon des critères généalogiques reste modeste. La définition de caractéristiques tumorales communes aux tumeurs du sein survenant dans un contexte de rédisposition lié à *BRCA2* a pour objectif l'identification de caractéristiques propres aux tumeurs BRCA2 permettant de mieux définir les indications de recherche de mutation de ce gène, et l'identification de facteurs impliqués dans la tumorigénèse des cancers du sein liés à *BRCA2*.

L'étude des profils génomiques des tumeurs BRCA2 a caractérisé la délétion récurrente des bras longs des chromosomes 13 et 14. L'analyse supervisée des données d'expression entre les tumeurs BRCA2 et les tumeurs familiales BRCA1 a identifié une signature spécifique des tumeurs BRCA2. Les exomes des chromosomes 13 & 14 pour 5 tumeurs informatives et leur ADN constitutionnel ont été séquencés afin d'identifier la ou les cibles des régions délétées. Cette analyse a permis la caractérisation de variants somatiques qui seront à étudier dans une large série de cas BRCA2 et contrôles pour conclure sur leur rôle dans la tumorigénèse liée à *BRCA2*.

La caractérisation de pertes de matériel chromosomique spécifiques aux tumeurs BRCA2, rapportée dans plusieurs études, offre une perspective diagnostique avec le développement d'un test FISH utilisable en pratique clinique pour préciser les indications d'une recherche de mutation du gène *BRCA2*, mais suggère également la présence de gènes cibles candidats dont l'inactivation est requise lors de la cancérisation mammaire liée à *BRCA2*.

**Mots clés :** Tumeurs du sein héréditaires, BRCA2, Signature transcriptionnelle, Profil génomique, NGS

**Laboratoire :** Institut Bergonié, INSERM U916 VINCO, 229 cours de l'Argonne, 33076 Bordeaux cedex

# ABSTRACT

---

## Gene expression and genomic profile study of BRCA2-mutant breast tumors

Germline *BRCA1* and *BRCA2* mutations account for 20-30% of familial breast cancer. The main indication for *BRCA2* screening is a family history, but the mutation detection rate in patients selected this way is low. The identification of characteristics common to BRCA2-associated tumors would improve the criteria used to select patients for *BRCA2* screening and could identify factors implicated in BRCA2-mutant breast cancer tumorigenesis.

The analysis of BRCA2-mutant breast tumor genomic profiles identified deletions of chromosomes 13q and 14q as a common feature of BRCA2-tumors. Supervised gene expression analysis of BRCA2-mutant breast tumors and familial breast tumors without germline *BRCA1* or *BRCA2* mutations identified a specific BRCA2 gene signature. Exome sequencing of chromosomes 13q and 14q for 5 BRCA2-mutant tumors, and their associated germline DNA was performed in order to identify the target(s) of the specific genomic deletions in the BRCA2 tumors. This analysis characterized somatic variants that will be screened for in a larger cohort of BRCA2 and control tumors cases to explore their role in BRCA2-mutant breast cancer.

Our study identified deletions of chromosomes 13q and 14q as a common feature of tumors with germline *BRCA2* mutations, as has been observed in several previous studies. We suggest that FISH analysis for the deletion of these chromosomes would be a rapid and technically feasible first step to select tumors worth screening for germline *BRCA2* mutations and we hypothesize that the inactivation of candidate genes located in these deleted regions allows the cell to resume division and progress thus contributing to tumorigenesis in BRCA2-mutant tumors.

**Keywords:** Hereditary breast cancer, BRCA2, Gene expression signature, Genomic profile, NGS

# TABLE DES MATIERES

---

Liste des Figures .....	7
Liste des Tables .....	8
Liste des Annexes .....	10
Abréviations .....	11
<b>1. INTRODUCTION .....</b>	<b>13</b>
<b>1.1. Altérations génétiques récurrentes dans les cancers du sein.....</b>	<b>13</b>
1.1.1. Réarrangements somatiques .....	13
1.1.2. Mutations `onctuelles somatiques.....	17
<b>1.2. Profil d'expression des cancers du sein.....</b>	<b>20</b>
1.2.1. Classification moléculaire des cancers du sein .....	20
1.2.2. Les nouveaux clusters intégratifs .....	24
<b>1.3. Prédiposition génétique au cancer du sein .....</b>	<b>27</b>
1.3.1. Les gènes de `rédi`osition au cancer du sein .....	27
1.3.2. Les gènes majeurs de `rédi`osition <i>BRCA1</i> et <i>BRCA2</i> .....	29
1.3.3. Caractéristiques des tumeurs du sein liées à une mutation constitutionnelle de <i>BRCA2</i> .....	31
<b>1.4. Fonctions cellulaires de la protéine <i>BRCA2</i> .....</b>	<b>33</b>
1.4.1. La `rotéine <i>BRCA2</i> : structure et transcri`tion.....	33
1.4.2. Rôles de la `rotéine <i>BRCA2</i> .....	35
<b>1.5. Développement d'une thérapie ciblée : le traitement anti<i>PARP</i> .....</b>	<b>37</b>
<b>2. OBJECTIFS DE L'ETUDE.....</b>	<b>40</b>
<b>3. RESULTATS.....</b>	<b>43</b>
<b>3.1. Caractérisation des échantillons de l'étude.....</b>	<b>43</b>
3.1.1. Criblage <i>BRCA1/BRCA2</i> .....	43
3.1.2. Po`ulation <i>BRCAX</i> .....	44
3.1.3. Caractéristiques anatomo`athologiques et cliniques.....	44
<b>3.2. Identification de profils génomique et transcriptomique spécifiques des tumeurs du sein liées à une mutation constitutionnelle <i>BRCA2</i> .....</b>	<b>46</b>
3.2.1. Etablissement d'une signature transcri`tomique des tumeurs <i>BRCA2</i> .....	46
3.2.2. Etude des `rofiles génomiques des tumeurs <i>BRCA2</i> .....	47
3.2.3. Mise en `lace d'un test <i>FISH</i> utilisable en routine ciblant les altérations génomiques associées aux tumeurs <i>BRCA2</i> .....	53
3.2.4. Publication .....	56

<b>3.3. Recherche d'un gène siégeant dans les régions délétées des chromosomes 13 et 14 dont l'inactivation serait nécessaire à la carcinogénèse mammaire liée à BRCA2.....</b>	<b>68</b>
3.3.1. Présentation de la technique NGS .....	68
3.3.2. Stratégie d'analyse des données NGS .....	77
<b>3.4. Variants identifiés et gènes candidats.....</b>	<b>89</b>
3.4.1. Confirmation des variants SNVs et INDELS identifiés.....	89
3.4.2. Hiérarchisation des variants SNVs et INDELS identifiés.....	93
3.4.3. Analyse bibliographique des gènes candidats en lien avec la tumorigénèse.....	95
<b>4. DISCUSSION ET PERSPECTIVES .....</b>	<b>100</b>
<b>4.1. Synthèse des résultats obtenus .....</b>	<b>100</b>
<b>4.2. Discussion et Perspective.....</b>	<b>103</b>
4.2.1. Application en pratique clinique des résultats obtenus .....	103
4.2.2. Signification biologique des remaniements génomiques .....	104
4.2.3. Limites de l'étude .....	107
4.2.4. Perspectives .....	110
<b>5. REFERENCES .....</b>	<b>112</b>
<b>6. ANNEXES .....</b>	<b>128</b>

# LISTE DES FIGURES

---

Figure 1: Classification hiérarchique supervisée de 115 relevements tumoraux et 7 tissus non malins selon la liste des « gènes intrinsèques » de Sorlie .....	22
Figure 2: Analyse de Kaplan-Meier sur deux cohortes de patients basée sur la classification des tumeurs du sein de Perou.....	24
Figure 3: Structure et protéines partenaires des protéines BRCA1 et BRCA2 .....	30
Figure 4 : Protéine BRCA2 : domaines, protéines partenaires, activités et fonctions.....	34
Figure 5 : Modèle de la fonction de BRCA2 dans la réparation des cassures double-brin par recombinaison homologue.....	37
Figure 6 : Représentation graphique de l'analyse d'un ADN tumoral sur un SNP .....	49
Figure 7 : Profil génomique des chromosomes 13 et 14 des tumeurs BRCA2 en BAC-CGH et en un SNP.....	51
Figure 8 : Test FISH susceptible d'identifier les tumeurs BRCA2 par délétion des régions d'intérêt des chromosomes 13 et 14 .....	55
Figure 9 : Extrait de l'analyse des données .fastq par le logiciel FastQC .....	70
Figure 10 : Résumé des différentes étapes bioinformatiques à partir de données de séquençage haut-débit.....	72
Figure 11: Extrait d'un fichier .fastq en codage Illumina 1.5+ issu d'un HiSeq2000.....	73
Figure 12 : Codage ASCII des scores de qualité QPhred en fonction des plateformes de séquençage .73	
Figure 13 : Extrait d'un fichier .sam .....	75
Figure 14 : Extrait d'un fichier .vcf.....	76
Figure 15 : Visualisation par IGV d'un fichier .bam.....	77
Figure 16 : Etude du variant non-sens c.1033G>T ; p.Glu345X du gène <i>PABPC3</i> sur la tumeur 149 et l'ADN constitutionnel associé .....	82
Figure 17 : Représentation graphique du log2 ratio des bras longs des chromosomes 13 et 14 obtenu par l'analyse des CNA à l'aide du package DNACopy et des données NGS de 5 tumeurs BRCA2 .....	87
Figure 18 : Etude de l'insertion d'une portion du bras long du chromosome 7 au niveau de l'exon 6 du gène <i>PCNX</i> localisé sur le chromosome 14 dans la tumeur BRCA2 133 .....	90
Figure 19 : Etude du remaniement génomique au niveau du gène <i>PCNX</i> en FISH sur la tumeur 133..92	
Figure 20 : Principe des algorithmes d'alignement de type « gapped » .....	109

# LISTE DES TABLES

---

Table 1 : Description des clusters intégratifs issus de l'analyse des données génomiques et transcriptomiques de près de 2000 cancers du sein.....	26
Table 2 : Distribution des caractéristiques différenciant les tumeurs du sein liées à BRCA2 des tumeurs du sein sporadiques.....	32
Table 3 : Description des mutations constitutionnelles du gène <i>BRCA2</i> des 9 échantillons BRCA2 de l'Institut Bergonié.....	44
Table 4 : Distribution des caractéristiques des 9 tumeurs du sein mutées pour <i>BRCA2</i> et des 36 tumeurs du sein BRCA2 de l'Institut Bergonié.....	45
Table 5 : Liste des gènes du cluster de sous-expression de 22 gènes de la signature BRCA2 localisés sur les chromosomes 13 et 14.....	47
Table 6 : Zone minimale de délétion des chromosomes 13 et 14 définie à partir de 4 profils CGH-BAC et 5 profils de puces SNP.....	52
Table 7 : Analyse bibliographique des bornes de délétion des chromosomes 13 et 14 des tumeurs BRCA2.....	52
Table 8 : Résultat de l'analyse des zones communes de délétion des chromosomes 13 et 14 sur 9 tumeurs BRCA2 et 9 tumeurs BRCA2.....	54
Table 9 : Présentation des 11 champs obligatoires du format <i>.sam Sequence Alignment/Map</i> en séquençage paired-end.....	74
Table 10 : Description des SNVs somatiques observés dans les tumeurs BRCA2 séquencées en NGS.....	80
Table 11 : Description des 10 SNVs somatiques étudiés en Sanger observés dans les tumeurs BRCA2 séquencées en NGS.....	82
Table 12 : Description des INDELS somatiques observées dans les tumeurs BRCA2 séquencées en NGS.....	83
Table 13 : Description des 6 INDELS somatiques étudiées en Sanger observées dans les tumeurs BRCA2 séquencées en NGS.....	84
Table 14 : Liste de gènes présents dans la tumeur à moins de 0.8 copie dans les zones minimales de délétion des chromosomes 13 et 14.....	88
Table 15 : Tableau récapitulatif des différentes altérations identifiées dans les tumeurs BRCA2 étudiées en NGS.....	93
Table 16 : Description des variants somatiques SNVs, INDELS et CNA de signification inconnue des 8 gènes candidats finaux.....	94



Table 17 : Gènes impliqués dans le cancer localisés sur la région minimale de délétion du chromosome 14.....105

# LISTE DES ANNEXES

---

Annexe 1 : Electro` h�rogramme des 10 variants SNVs somatiques exoniques non iso-s�mantiques s�quenc�s sur la tumeur et l'ADN contitutionnel corres` ondant.....	128
Annexe 2 : Electro` h�rogramme des 4 variants INDELS somatiques exoniques s�quenc�s sur la tumeur et l'ADN constitutionnel corres` ondant .....	130
Annexe 3: Re` r�sention de la visualisation des fichiers d'alignement .bam ` ar l'outil IGV sur la tumeur et l'ADN constitutionnel corres` ondant. ....	131
Annexe 4 : Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations.....	133

# ABBREVIATIONS

---

AA	Acide Aminé	LRR	<i>Log R ratio</i>
ADN	Acide désoxyribonucléique	miRNA	Micro-ARN
ARN	Acide ribonucléique	NGS	<i>Next Generation Sequencing</i>
BAC	<i>Bacterial Artificial Chromosome</i>	NLS	<i>Nuclear Localisation Signal</i>
BAF	<i>B Allele Frequency</i>	NT	Nucléotides
bam	Version binarisée du format sam	PARP1	Poly(ADP-ribose) polymérase 1
BRCA1	<i>Breast Cancer 1</i>	pb	Paire de bases
BRCA2	<i>Breast Cancer 2</i>	QMPSF	<i>Quantitative multiplex PCR on short fragments</i>
BRCT	BRCA1 C-Terminus	RA	Récepteur aux androgènes
CGH	<i>Comparative Genomic Hybridization</i>	RE	Récepteur aux œstrogènes
CNA	<i>Copy Number Aberrations</i>	RH	Récepteurs hormonaux
EMMA	<i>Enhanced Mismatch Mutation Analysis</i>	RMA	<i>Robust Multi-array Analysis</i>
FISH	<i>Fluorescence in situ hybridization</i>	RP	Récepteur à la progestérone
GSEA	<i>Gene Set Enrichment Analysis</i>	sam	<i>Sequence Alignment/Map</i>
INDEL	<i>Insertion-délétion</i>	SNP	<i>Single Nucleotide Polymorphism</i>
kb	Kilobase	SNV	<i>Single Nucleotide Variant</i>
limma	<i>Linear Models for Microarray Data</i>	vcf	<i>Variant call format</i>
LOH	<i>Loss Of Heterozygosity</i>		

# INTRODUCTION

---

# 1. INTRODUCTION

---

## 1.1. Altérations génétiques récurrentes dans les cancers du sein

L'hypothèse de Boveri en 1914 selon laquelle le cancer serait lié à l'accumulation d'altérations génomiques [1] a été confortée en 1960 lors de la description du chromosome de Philadelphie par Nowell et Hungerford dans la leucémie myéloïde chronique [2]. L'évolution des connaissances a permis d'aboutir à un consensus décrivant le cancer comme une maladie génétique liée à deux types majeurs d'évènements: l'inactivation de gènes suppresseurs de tumeurs par délétion, mutation ou mécanisme épigénétique et l'activation ou la dérégulation d'oncogènes par mutation ponctuelle, amplification ou translocation [3].

### 1.1.1. Réarrangements somatiques

#### 1.1.1.1. Altération du nombre de copies

La cytogénétique traditionnelle a très vite mis l'accent sur les aneusomies qui caractérisaient le caryotype des cellules tumorales. En ce qui concerne le caryotype du sein, trois remaniements majeurs ont été caractérisés : la translocation t(1;16), l'isochromosome 1q et les amplifications géniques [4]. L'analyse du génome des cancers du sein par hybridation génomique comparative (CGH) a permis de préciser ces anomalies. Sur une cohorte de 305 tumeurs du sein, l'équipe de Rennstam identifie des gains de régions chromosomiques dans 270 cas et des pertes dans 216 cas [5].

Selon diverses études, les régions délétées communes à plusieurs tumeurs mammaires concernent les bras chromosomiques 1p, 8p, 13q, 16q et les régions régulièrement gagnées impliquent les bras chromosomiques 1q, 8q, 17q et 20q [5-9].

L'avancée technologique dans l'analyse des cartes de matériel par la CGH hautement résolutive, par l'analyse de carte d'hétérozygotie (ou *LOH* pour *Loss Of Heterozygosity*) sur carte SNP (*Single Nucleotide Polymorphism*) et enfin par le séquençage nouvelle génération (NGS pour *Next Generation Sequencing*), permet d'étudier sur l'ensemble du génome tumoral les régions délétées mais aussi les réarrangements complexes nommés réarrangements de grande taille entraînant la délétion ciblée d'un ou plusieurs exons d'un gène. En 2009, l'analyse de 24 tumeurs du sein (tumeurs sporadiques et héréditaires, et lignées de cancers du sein) en séquençage haut-débit par l'équipe de Stratton a révélé 2166 réarrangements somatiques dont 357 délétions, soit 16.5% des réarrangements somatiques isolés [10]. Une récente étude associant données génomiques et transcriptomiques sur près de 2000 cancers du sein, a mis en exergue trois régions génomiques délétées de manière récurrente, associées à une sous-expression des gènes à proximité de ces régions en 8<sup>12</sup> (*PPP2R2A*- *protein phosphatase 2, regulatory subunit B, alpha*), en 9<sup>21</sup> (*MTA*- *methylthioadenosine phosphorylase P*) et en 17<sup>11</sup> (*MAP2K*- *mitogen-activated protein kinase kinase 4*), suggérant un rôle suspecté de tumeur des gènes concernés [11].

L'amplification est une altération génétique fréquemment observée dans le cancer du sein [12]. Cette altération est décrite dans 28.5% des échantillons de l'étude de Stratton [10]. Les études de cytogénétique ont montré qu'il existait des chromosomes double minute, fragment chromosomique de très petite taille sans centromère, et des régions *homogeneously staining regions* (HSR), régions de taille variable présentes au sein d'un ou plusieurs chromosomes, corrélées à des amplifications d'oncogènes. Les régions les plus fréquemment amplifiées sont les régions 1q31-q32, 8<sup>12</sup>, 8q24, 11q13, 12<sup>13</sup>-ter, 16<sup>13</sup>, 17q12, 17q22-q24 et 20q13 [13-16]. Les amplifications les plus fréquentes ciblant des oncogènes connus concernent les régions suivantes : 8<sup>12</sup> *FGFR1* (*fibroblast growth factor receptor 1*) et *ZNF703* (*zinc finger protein 703*); 8q24 *MYC* (*myelocytomatosis viral oncogene*); 11q13 *CCND1* (*Cycline D1*); 17q12 *ERBB2* (*V-Erb-B2 avian erythroblastic leukemia viral oncogene homolog 2*). Les cibles potentielles de ces amplifications concernent les gènes suivants : *MED24* (*mediator complex subunit 24*) et *GRB7* (*growth factor receptor-bound protein 7*) localisés sur l'amplicon 17q12 ; *RPS6KBI* (*ribosomal protein S6 kinase polypeptide 1*), *PATI* (*protein interacting with amyloid precursor protein tail 1*) et *TBX2* (*T-box 2*) localisés sur l'amplicon 17q22-q24 ; les gènes *ZNF217* (*zinc finger protein 217*), *RAE1* (*RNA export 1*), *NCOA3* (*nuclear receptor coactivator 3*) et *CYP24A1* (*Cytochrome P450, Family 24, Subfamily A, Polypeptide1*) localisés sur l'amplicon 20q13.

De récentes études suggèrent que les amplifications de régions normalement ségréguées dans le génome ne sont pas des événements indépendants. Par-exemple, la co-amplification des régions 8<sup>11-8<sup>12</sup></sup> et 11q12-11q14 [17] ou 8q24 et 17q21 [18] a été rapportée comme impliquée dans la physiopathologie du cancer du sein et suggère une coopération entre les gènes localisés sur chaque amplicon lors de

l'oncogenèse [19]. Les amplicons sont des régions composites qui englobent des régions génomiques d'origines différentes réarrangées et amplifiées au sein d'un même amplicon.

#### 1.1.1.2. Translocations

Les réarrangements somatiques peuvent entraîner la transformation de deux gènes en un oncogène. Certaines anomalies sont spécifiques d'une tumeur maligne et leurs présences peuvent avoir un intérêt diagnostique, thérapeutique et pronostic [20]. La translocation dans la cancérogénèse met en évidence deux mécanismes majeurs : (i) la dérégulation de l'expression de gènes, suite à la translocation d'un gène sous le contrôle d'un promoteur fort (ii) et la formation de gènes de fusion, suite à la translocation de deux gènes et de l'expression d'une protéine chimère contenant des éléments de séquence des deux gènes. A ce jour, en-dehors des tumeurs du sein rares, aucune translocation récurrente aboutissant à la dérégulation de l'expression d'un gène n'a été décrite dans le cancer du sein.

Il est à noter que l'analyse NGS de l'équipe de Stratton informe sur la prédominance des réarrangements intrachromosomiques (1311 contre 239 interchromosomiques) dans les cancers du sein. Cette donnée n'était pas documentée auparavant en raison de la résolution limitante des techniques de cytogénétique.

Il est difficile de lister de manière exhaustive les translocations décrites dans le carcinome mammaire en raison de l'hétérogénéité des translocations et de l'identification de nouvelles translocations régulièrement. L'équipe de Mitelman a mis en place une base de données recensant un nombre important des anomalies chromosomiques sur tous les types tumoraux dont le cancer du sein (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>) [21]. Il existe cependant une translocation récurrente particulièrement observée dans les carcinomes de bas grade, exprimant le Récepteur aux Œstrogènes (RE), impliquant les chromosomes 1 et 16 et aboutissant à la formation d'un dérivé chromosomique der(1;16)(q10;q10) [22-27]. Cette translocation déséquilibrée entraîne la perte du bras long du chromosome 16 et le gain du bras long du chromosome 1. Des études de *LOH* ont identifié trois régions minimales de délétion (16q22.1, 16q23.2-24.1 et 16q24.3) suggérant la présence d'un gène suppresseur de tumeur sur le bras long du chromosome 16 [28]. Un gène candidat dans la région 16q24.3 est le gène *CDH15* (*M-Cadherin*) [29]. Le gène *CTCF* (*CCCTC-binding factor*), régulateur négatif de l'oncogène *MYC*, localisé en 16q22.1 est également candidat [30].

Sur cette même région, un autre gène codant pour une cadhérine, *CDH1* (*E-Cadherin*) localisé en 16q22.1, jouerait un rôle dans la cancérogénèse des tumeurs lobulaires [31]. *CDH1* est spécifiquement muté dans les carcinomes lobulaires infiltrants sur un allèle tandis que le second allèle est délété, mais

les délétions du 16q sont également observées dans les carcinomes canaux infiltrants, suggérant que la cible de cette délétion non encore identifiée est autre que *CDH1*.

Les carcinomes mammaires, peuvent présenter des gènes de fusion à une fréquence très faible inférieure à 1% [32]. Le carcinome mammaire sécrétant, entité histologique très rare du sujet jeune, présente une translocation t(12;15) impliquant le gène de fusion *ETV6-NTRK3* [33]. Cette translocation conduit à la fusion entre l'exon 5 du gène *ETV6* (*ETS family transcription factor variant 6*) et l'exon 13 du gène *NTRK3* (*TRK family tyrosine kinase receptor for neurotrophin-3*), et à l'exclusion de l'exon 16 de *NTRK3* codant pour une séquence régulant négativement l'activité tyrosine kinase de *NTRK3*. La protéine chimérique exprime une activation constitutive du domaine tyrosine kinase sous la dépendance du domaine de dimérisation du facteur de transcription *ETV6* indépendante du ligand [34].

Le carcinome adénoïde kystique, autre type d'articulier de carcinome mammaire, se caractérise par la translocation équilibrée t(6;9) induisant la fusion des gènes *MYB* (*myeloblastosis viral oncogene*) et *NFIB* (*nuclear factor I/B*) [35].

Une autre translocation récurrente a été identifiée plus particulièrement dans les tumeurs triplonégatives (cf. paragraphe 1.2.1), concernant les gènes *MAGI3* (*membraneassociated guanylate kinase, WW and PDZ domain containing 3*) sur le chromosome 17 et *AKT3* (*v-akt murine thymoma viral oncogene homologue 3*) sur le chromosome 1q, aboutissant à une translocation équilibrée de l'intron 9 de *MAGI3* à l'intron 1 de *AKT3* avec perte du second domaine PDZ (nomenclature correspondant à l'acronyme des trois premières protéines découvertes qui étaient pourvues de ce motif de 90 à 100 acides aminés : PSD-95 -*Post Synaptic Density de 95 kDa*-, DLG-*Drosophila melanogaster disc large protein*-, ZO-1 -*zonula occludens-1 protein*) capable entre autre de lier *PTEN* (*phosphatase and tensin homolog*) et nécessaire à la régulation négative de ce dernier dans la voie phosphatidylinositol-3-kinase (PI3K) [36]. La fusion *MAGI3-AKT3* entraîne l'activation constitutive de la kinase *AKT* [36]. Dans la lignée cellulaire MDAMB-175, une translocation entre le gène *NRG1* (*neuregulin-1*) situé en 8p12 et le gène *TENM4* (*Teneurin Transmembrane Protein 4*) localisé en 11q14, conduit à la formation d'un gène de fusion nommé *γ-Heregulin* dont la partie 5' est constituée du gène *DOC-4* et la partie 3' du gène *NRG1* [37]. Le gène de fusion sous le contrôle du promoteur de *DOC-4*, code pour un facteur de croissance autocrine [38].

Vingt-deux autres gènes de fusion ont été décrits dans l'étude du génome de 24 cancers du sein de Stratton, en particulier le gène *ETV6* dont plusieurs partenaires de fusion ont été identifiés dans différents types tumoraux [39]. Un réarrangement intrachromosomique sur le bras court du chromosome 12 conduit à la fusion entre les exons 1 et 2 du gène *ETV6* et des exons 35 à 57 du gène *ITPR2* (*inositol 1,4,5- trisphosphate receptor, type 2*).



### 1.1.2. Mutations somatiques

Les mutations somatiques (substitution nucléotidique ou petite insertion-délétion) représentent un mécanisme su`élémentaire de la tumorigénèse [19]. La multiplication des études en NGS sur diverses cohortes de carcinomes mammaires a permis d'explorer les mutations somatiques dans les cancers du sein et d'identifier de nouveaux gènes impliqués dans la tumorigénèse mammaire. L'étude menée par l'équipe de Meyerson publiée dans Nature en 2012, a identifié 6 gènes fréquemment mutés dans 103 cancers du sein [36]. L'un d'eux, le gène *CBFB* (*Core-Binding Factor, Beta Subunit*), impliqué dans l'inversion du chromosome 16 avec le gène *MYM11* dans les leucémies aiguës myéloïdes de type M4Eo [40] est rapporté pour la première fois comme gène fréquemment muté dans le cancer du sein et autres cancers épithéliaux ; les cinq autres gènes ont quant à eux précédemment été identifiés comme gènes impliqués dans la tumorigénèse mammaire. Il s'agit des gènes *TP53* (*Tumor protein p53*) [41], *PIK3CA* (*Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic Subunit Alpha*) [42], *AKT1* (*Akt Murine Thymoma Viral Oncogene Homolog 1*) [43], *GATA3* (*GATA binding protein 3*) [44] [36;44;44], et *MAP3K1* (*MAP kinase kinase kinase 1, E3 (ubiquitin protein ligase)*) [36;45]. Dans le cancer du sein, la fréquence de mutation de ces gènes varie selon les sous-types tumoraux. Cependant, au sein d'une cohorte de cancers mammaires, les mutations somatiques des gènes *TP53* et *PIK3CA* sont observées dans environ 30% des cas, les gènes *AKT*, *GATA3*, *CBFB* et *MAP3K1* sont mutés dans moins de 5% des cas [36;45;46].

L'étude de l'exome de 100 cas de cancers du sein par l'équipe de Stratton en 2012 a identifié 7241 mutations somatiques dont 277 insertions-délétions [46].

### 1.1.3. Anomalies épigénétiques

L'épigénétique est définie par l'étude des changements dans l'expression des gènes qui sont maintenus lors de la division cellulaire, et qui ne résultent pas de modifications de la séquence de l'ADN [47]. Les mécanismes principaux de l'épigénétique sont la méthylation de l'ADN, la modification des histones, l'expression de microARN (miRNA). Ces modifications épigénétiques affectent l'expression des gènes.

La modification du domaine N-terminal des histones incluant l'acétylation, la méthylation, l'ubiquitination, la phosphorylation et la ribosylation peut significativement affecter l'expression des gènes en modifiant la topologie spatiale du nucléosome avec une conformation plus ou moins relâchée de la chromatine [48]. Dans le cancer du sein, il a été montré que la modification d'histone, en association avec la méthylation de l'ADN est fréquemment associée à la répression de gènes su`resseurs de tumeurs. La répression inappropriée de ces gènes favorise la tumorigénèse. La

méthylation de la queue des histones est assurée par des histones méthyl-transférases (HMT), la méthylation des histones H3 lysine 4 (H3K4), H3K36 et H3K79 active la transcription, contrairement à la méthylation des H3K9, H3K27 et H4K20 qui a un effet répressif. L'action coordonnée de deux HMT, le complexe Polycomb (PcG) et les protéines Trithorax, régule la structure dynamique de la chromatine.

L'acétylation dépend de l'activité des histones acétyltransférases (HAT), qui agissent comme des coactivateurs de la transcription et des histones déacétylases (HDAC), associées à une répression de l'activité transcriptionnelle.

La méthylation de l'ADN est régie par les méthyl-transférase d'ADN (DNMT) qui ont la capacité d'ajouter un groupement méthyle au niveau des résidus cytosine contenus dans les îlots CpG. La méthylation de l'ADN est impliquée dans de nombreuses fonctions biologiques comme l'inactivation d'un des chromosomes X chez la femme et l'empreinte parentale.

L'hyperméthylation de l'ADN au niveau des régions non régulatrices et d'éléments structuraux comme l'ADN centromérique augmente l'instabilité chromosomique et favorise la tumorigénèse [49-51]. L'hyperméthylation dans les régions promotrices de gènes en favorise l'expression. Ainsi la perte de méthylation d'un oncogène entraînera sa surexpression et facilitera le processus tumoral, comme les oncogènes *RAS* et *MYC* identifiés comme étant hypométhylés dans les tissus cancéreux [52-54].

L'hyperméthylation de l'ADN est associée à l'inhibition de la transcription de gènes et indirectement à l'instabilité génomique lorsqu'elle cible des gènes suppresseurs de tumeurs [55;56]. Un certain nombre de gènes participant aux différentes étapes du développement tumoral sont hyperméthylés et réprimés dans les cancers avec une spécificité pour certains types de tumeurs [57]. Ainsi le second événement conduisant à l'inactivation d'un gène suppresseur de tumeur peut être lié à une mutation épigénétique comme la méthylation de son promoteur. Dans le cancer du sein, on retrouve entre autre la méthylation du gène *TSM1* (*target of methylation-induced silencing 1*) codant pour une caspase pro-apoptique, du gène de réparation de l'ADN *BRCA1* (*breast cancer 1*), du gène suppresseur de tumeur pro-apoptique *RASSF1A* (*Ras association domain family 1*) mais aussi des gènes codants pour les récepteurs hormonaux RE et RP (Récepteur à la Progestérone) [58;59].

Ces modifications épigénétiques constituent une nouvelle approche thérapeutique. Il existe deux classes d'inhibiteurs de DNMT (DNMTi) : les analogues nucléosidiques et les analogues non-nucléosidiques. Les analogues nucléosidiques tels que la 5-azacytidine et la 5-aza-2'-désoxycytidine (décitabine) piègent les DNMT en s'intégrant dans l'ADN à la place de résidus cytosine. La décitabine est actuellement utilisée dans le traitement des leucémies aiguës myéloïdes. Les DNMTi analogues non-nucléosidiques inhibent la méthylation de l'ADN en se liant directement au domaine catalytique de l'enzyme DNMT sans s'incorporer dans l'ADN. Pour exemple, la procaine et son dérivé

Les inhibiteurs de la méthyltransférase ont montré une action déméthylante conduisant à la réexpression de gènes suppresseurs de tumeur réprimés par l'hyperméthylation [60].

Les inhibiteurs d'HDAC ont un mécanisme d'action mal connu à ce jour. Ces molécules induisent dans les cellules tumorales l'arrêt du cycle cellulaire, la différenciation, et l'apoptose. De plus, ces agents ont une action anti-angiogénique. L'action des inhibiteurs d'HDAC est médiée par la réactivation de l'expression du gène *p21* (*cyclin-dependent kinase inhibitor 1A*) impliqué dans les points de contrôle des phases G1 et G2 du cycle cellulaire ainsi que dans la différenciation [61]. Actuellement, le vorinostat, inhibiteur d'HDAC, est utilisé dans la prise en charge des lymphomes cutanés à cellules T.

La surexpression de la protéine EZH2 (*enhancer of zeste homolog 2*) appartenant au complexe répressif Polycomb (PRC) a été observée dans les cancers du sein et du poumon, et a été associée à un marqueur d'agressivité [62]. Cette protéine catalyse l'addition de groupes méthyle sur H3K27. La réactivation des gènes réprimés par le PRC par l'action de l'inhibiteur de EZH2, le 3-Deazaneplanocin A (DZNep), entraîne l'apoptose des cellules tumorales. Cette action pro-apoptotique serait médiée par le gène *FBX032* qui code pour la protéine MAFBx (*muscle atrophy F-box protein*). L'expression de cette protéine pourrait être modulée par la voie PI3K (*Phosphoinositide 3-Kinases*)/Akt et serait diminuée dans les tumeurs primitives du sein [63;64].

Enfin le dérèglement de miRNA (microARN) a récemment été associé aux cancers. Les miRNA sont des molécules d'ARN de 18 à 23 nucléotides codées par le génome. Ce sont des régulateurs post-transcriptionnels capables de contrôler l'expression de gènes. Après la transcription assurée par la polymérase II, le pré-miRNA subit plusieurs étapes de maturation à l'aide des protéines Drosha, Dicer et le complexe protéique RISC (*RNA-Induced Silencing Complex*) [47]. Lorsque le miRNA se lie à son ARN cible, si la complémentarité est totale, l'ARN cible est dégradé et si la complémentarité est partielle, il bloque la traduction de l'ARN cible.

Le rôle des miRNA dans l'initiation et la progression tumorale a clairement été établi ces dernières années. En particulier, le rôle métastatique des miRNA a été démontré en premier par Ma et son équipe du groupe de Weinberg, lors de la découverte de l'implication du *miR-10b* dans l'invasion tumorale et la formation de métastases dans le cancer du sein [65]. Ainsi, la dérégulation de plusieurs miRNA a été décrite dans les cancers du sein, comme les miRNA oncogènes *miR-10b* (régule *Homeobox D10*), *miR-21* (régule *TPM1-Tumor suppressor tropomyosin 1*, *PDCD4-programmed cell death 4*, *maspin-mammary serine protease inhibitor* et *PTEN*), le cluster *miR-17~92* (régule *E2F1-E2F transcription factor 1*, *Bim*) et les miRNA suppresseurs de tumeurs *miR-145* (régule *MYC*), la famille *miR-200* (régule *TGFβ2 - Transforming growth factor beta*, *PLCG1 - Phospholipase C, Gamma 1*) et la famille *let-7* (régule *HRAS (Harvey Rat Sarcoma Viral Oncogene Homolog)* et *HMGA2-high mobility group AT-hook 2*) [66].

## 1.2. Profil d'expression des cancers du sein

### 1.2.1. Classification moléculaire des cancers du sein

L'hétérogénéité des cancers du sein rend 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'adoption d'outils regroupés sous le terme générique de bio-informatique. Ces techniques permettent l'analyse du génome à grande échelle ainsi que l'analyse du transcriptome par la quantification de l'expression des gènes. Ces analyses sur bio-informatique 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 [67] est que la diversité génétique des cancers du sein peut être expliquée par une diversité au niveau de l'expression des gènes, étudiée par micro-array. C'est à cette époque que l'on doit une première classification moléculaire des cancers du sein en 4 sous-types sur des données transcriptomiques de 65 tumeurs [67], affinée à l'aide d'études successives en 2001, 2003 et 2004 [68-70] sur un plus grand nombre de prélèvements, et définissant les 5 sous-types suivants : Basal-like, ERBB2+, Normal-Breast-like, Luminal B et Luminal A (Figure 1). Des données pronostiques ainsi que l'étude du taux mutationnel de gènes associés au cancer du sein en fonction des 5 sous-types ont également été réalisées [68;71].

Ces sous-types se répartissent en deux grands groupes : un groupe caractérisé par une expression faible ou nulle du RE codé par *ESR1* (estrogen receptor 1), tumeurs dites « RE négatives » et l'autre par une expression plus ou moins forte de celui-ci, tumeurs dites « RE positives ».

Le groupe des « RE négatives » comprend (i) les tumeurs dites basales, correspondant le plus souvent aux tumeurs dites « triple négatives » après analyse des récepteurs RE, RP et de l'expression de HER2 codée par *ERBB2* (*Erythroblastic Leukemia Viral Oncogene Homolog 2*) par immunohistochimie (absence de marquage RE et RP et absence de surexpression de HER2) ; (ii) les tumeurs ERBB2 positives (par amplification du gène *ERBB2*) ; (iii) les tumeurs dites « normales like ».

Les cancers « basal-like » sont de haut grade histologique, l'index mitotique est souvent élevé [72;73]. Ce sous-type est caractérisé par l'absence d'expression du cluster Luminal et une négativité d'expression des récepteurs hormonaux RE et RP, et de l'absence de surexpression de HER2 (est dit « triple négatif »). Ces tumeurs expriment fortement des gènes d'origine myoéithéliale, 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 17*), et *EGFR* (*Epidermal Growth Factor Receptor*) [74-76]. Les gènes ex`rimés `ar ces tumeurs sont majoritairement impliqués dans le contrôle du cycle cellulaire [76]. Ce groupe `résente à la fois le `lus haut taux de mutation du gène *TP53* (80%) et le `lus bas du gène *PIK3CA* (9%). A noter la fréquence de 5% de mutations du gène *MLL3* (*mixed-lineage leukemia 3*), taux semblable aux autres groupes [71]. Ce gène code `our une `rotéine appartenant au complexe ASCOM (*ASC-2- and Mll3-containing complex*) qui `ossède une activité histone méthyl-transférase. Ce gène est décrit comme `otentiel gène su`resseur de tumeur en raison de la fréquence des mutations dans divers types tumoraux et l`a`arition de tumeurs lors de son inactivation chez des souris [77].

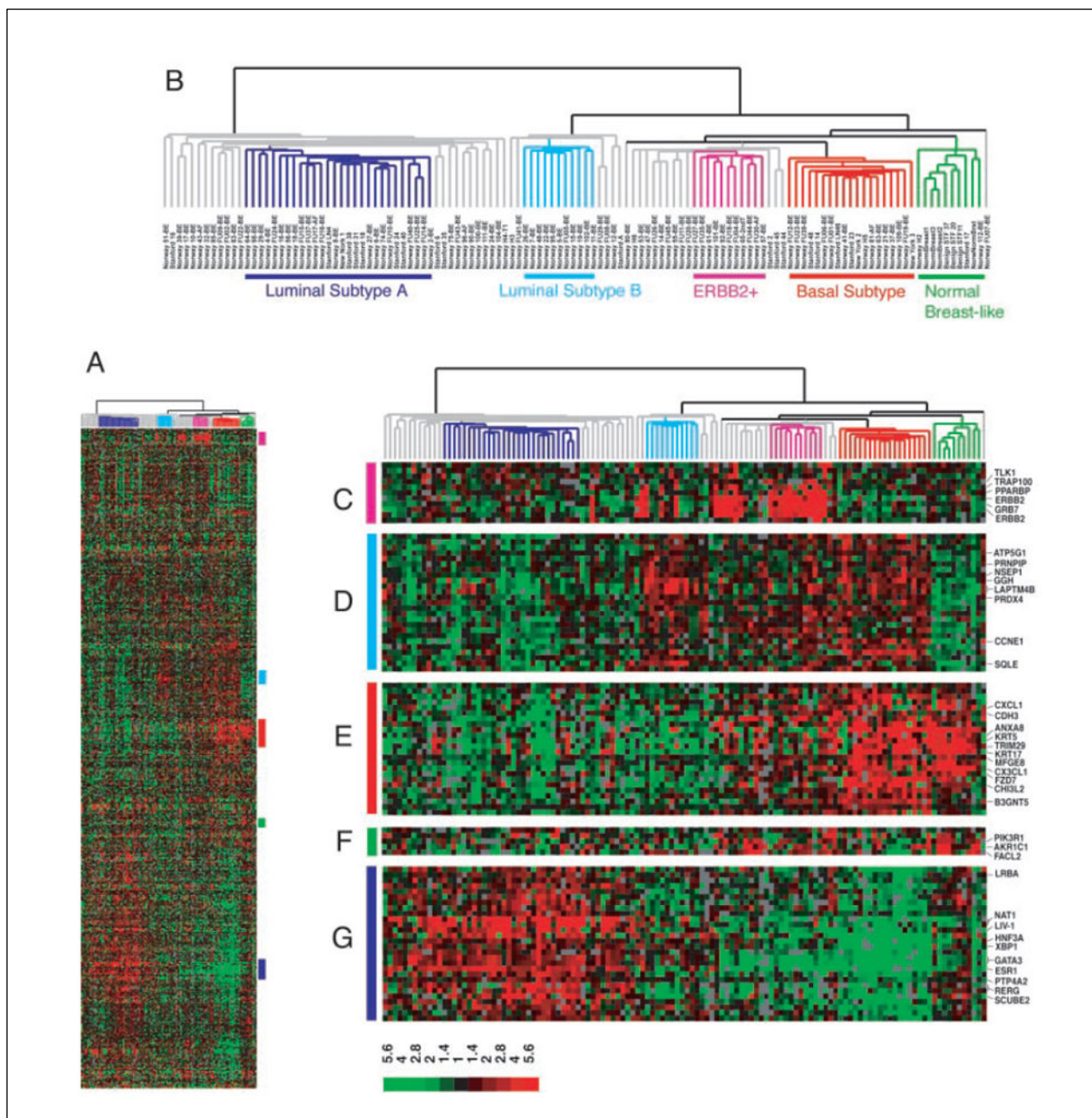
Les cancers « ERBB2+ » sont caractérisés `ar une surex`ression du gène *ERBB2* et des autres gènes situés dans l`am`licon 17q22.24 comme *MED24* (*mediator complex subunit*) et *GRB7* (*growth factor receptor-bound protein 7*) [69]. Ces tumeurs n'ex`riment `as les réce`teurs hormonaux RE et RP. Ces tumeurs `résentent un taux de mutation de *TP53* s`élevant à 72% et de *PIK3CA* égal à 39% [71].

Enfin, les cancers « Normal Breast-like », sont caractérisés `ar des gènes connus `our être ex`rimés dans le tissu adi`eux et les cellules non é`ithéliales.

Le groupe des « RE `ositives » rassemble les tumeurs lumineales A et B.

Les tumeurs lumineales A sont des cancers de bas grade `résentant une faible `rolifération et une forte ex`ression de *ESR1*, *GATA3* (*GATA binding protein 3*), *HNF3α* (*Hepatocyte Nuclear Factor 3α*), *XBPI* (*X-box binding protein 1*) et *LIV-1* (*oestrogen regulated protein*) [69]. Elles `résentent un taux de mutation de *TP53* d`environ 13% [68]. Ces tumeurs ex`riment majoritairement des gènes impliqués dans le métabolisme des acides gras et la voie de signalisation des réce`teurs aux œstrogènes [69;70]. Le nombre de gènes fréquemment mutés s`avère être le `lus im`ortant dans ce sous-groupe avec les gènes *PIK3CA* (45%), *GATA3* (14%), *MAP3K1* (*MAP kinase kinase kinase 3*) (13%), *TP53* (12%), *CDH1* (9%), *MLL3* (8%) et *MAP2K4* (7%) [71]. Les mutations inactivatrices de *MAP3K1* et *MAP2K4* sont mutuellement exclusives. Ces deux gènes sont des acteurs de la voie de stress `38-JNK1 (*Jun N-terminal Kinase (JNK) 1*).

Les tumeurs lumineales B sont de `lus haut grade, ont un index de `rolifération `lus im`ortant et ex`riment en moindre quantité les réce`teurs hormonaux RE et RP. Elles sont caractérisées `ar une moindre ex`ression des gènes ex`rimés `ar les tumeurs du groupe « Luminal A », et `ar une forte ex`ression des gènes impliqués dans le cycle cellulaire tels que les gènes *CCNE1* et *CCND1* (*Cycline E1 et D1*), mais aussi les gènes *GGH* (*γGlutamyl Hydrolase Precursor*), *NSEPI* (*Nuclease Sensitive Element Binding Protein 1*), *LAPTM4β* (*lysosomal protein transmembrane 4 β*) et *MYB* (*myeloblastosis viral oncogene homolog*) [69;75]. Les tumeurs lumineales B `résentent une mutation des gènes *PIK3CA* et *TP53* à une fréquence de 29%, une mutation du gène *GATA3* à 15%, et une fréquence faible de mutation des gènes *MAP3K1* (5%), *CDH1* (5%) et *MLL3* (6%) [71].



**Figure 1: Classification hiérarchique supervisée de 115 prélèvements tumoraux et 7 tissus non malins selon la liste des « gènes intrinsèques » de Sorlie (Extrait de Sorlie et al 2003) [70]. (A) Représentation du cluster entier des 534 gènes et 122 échantillons, (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, « Basal-like » en orange et « Normal Breast-like » en vert.**

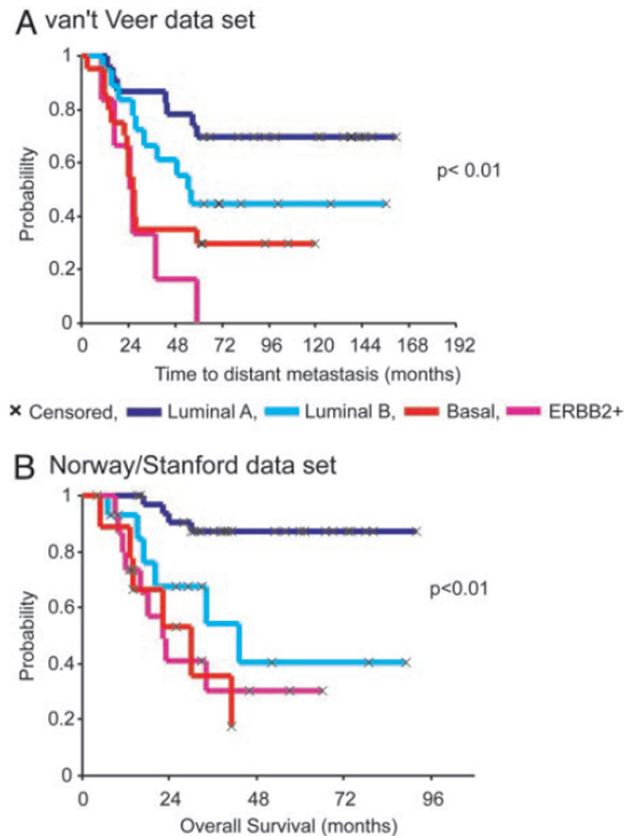
Les cinq groupes moléculaires identifiés ont été corrélés à des différences de survie et de réponse au traitement. Les analyses des courbes de survie de Kaplan-Meier permettent d'associer les sous-types « basal-like » et « ERBB2+ » au plus mauvais pronostic. Le sous-type « Luminal B » présente également un moins bon pronostic que les tumeurs « Luminal A » [70] (Figure 2). La similarité du

Le 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 a apporté des valeurs pronostiques d'estimation du risque de rechute ou du bénéfice d'un traitement spécifique [78;79].

Suite à l'identification initiale de 5 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 andocrine » [80], qui partage les caractéristiques du sous-groupe « ERBB2+ » et semble représenter une activation de la voie de signalisation du récepteur aux androgènes (RA) ; (ii) le sous-type « interferon » [75] 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 » [81] qui inclut les tumeurs ayant un phénotype « stem cell-like ».

De plus, la classification non supervisée de Farmer *et al.*, identifie trois groupes de tumeurs distincts [80]. Le groupe des tumeurs RE+ RA+ RP+ qui rassemble les tumeurs lumorales, le groupe RE- RA- RP- qui contient les tumeurs basales et enfin le groupe RE- RA+ qui contiendrait des tumeurs provenant de tissus engagés dans une différenciation androcrine. Ces tumeurs sont confondues avec les tumeurs ERBB2 car elles présentent fréquemment l'amplification 17q.

Parallèlement à l'identification des sous-types tumoraux, le profil d'expression des gènes a été étudié par plusieurs équipes dans un but pronostic. Le Netherlands Cancer Institute d'Amsterdam a établi la signature moléculaire MammaPrint® qui tient compte de l'expression de 70 gènes, permettant de prédire l'apparition de métastases à 5 ans d'une tumeur sans atteinte ganglionnaire [82]. Une étude pronostique sur des cancers du sein RE+ localisés, sans envahissement ganglionnaire et traités par Tamoxifène, a établi un score de rechute basé sur l'étude en RT (*reverse transcriptase*)-PCR de 21 gènes appelé OncotypeDX [83]. Ces deux signatures sont actuellement testées dans des essais cliniques randomisés.



**Figure 2: Analyse de Kaplan-Meier sur deux cohortes de patients basée sur la classification des tumeurs du sein de Perou (Extrait de Sorlie et al, 2003) [70]. (A) Survie sans rechute sur 97 cas de cancers du sein sporadiques de Van't Veer et al, 2002 [82], (B) Survie générale de 72 patients de la cohorte Norway/Stanford. Les courbes sont colorées en fonction du sous-type moléculaire : « Luminal A » en violet, « Luminal B » en bleu, « ERBB2+ » en rose, « Basal » en orange.**

### 1.2.2. Les nouveaux clusters intégratifs

L'analyse conjointe des données génomiques et transcriptomiques, associée à une analyse statistique sur près de 2000 cancers du sein, a abouti à la définition de nouveaux sous-groupes biologiques appelés « clusters intégratifs » [11]. Le *clustering* intégratif permet d'identifier 10 classes différentes de cancers du sein (IntClust1 à 10) corrélés ou non à la classification intrinsèque. Ils sont définis par des altérations génomiques spécifiques et des valeurs pronostiques distinctes (Table 1).

Sept clusters intégratifs contiennent plus de 50% de tumeurs lumineales. L'un d'eux, le IntClust2, caractérisé par une amplification en 11q13 et/ou en 11q14, présente le plus mauvais pronostic des clusters ER+. L'analyse de l'expression des gènes de ce cluster, montre un enrichissement en gènes impliqués dans le cycle cellulaire, et tout particulièrement dans le passage de la phase G1 à la phase S comme *CCND1*.



Les clusters 3, 7 et 8 rassemblent des tumeurs lumineales A de bas grade dans leur majorité, et sont de bon pronostic. Les tumeurs du cluster 3, de profil génomique lat, présentent le plus haut taux de mutation des gènes *PIK3CA* (≈55%) et *CDH1* (>20%) ainsi que le meilleur pronostic des clusters lumineux (survie à 10 ans supérieure à 90%). Le profil génomique du IntClust7 est caractérisé par une perte en 16q, un gain en 16 et une amplification en 8q. De plus, le taux de mutation du gène *MAP3K1* y est le plus important (>20%). Enfin, la translocation t(1;16) récemment décrite dans les tumeurs ER+ de bas grade (cf. paragraphe 1.1.1.2) caractérise les tumeurs du cluster 8. Un fort taux de mutation des gènes *PIK3CA*, *GATA3* et *MAP2K4* est également associé à ce groupe.

Les clusters 1, 6 et 9 sont constitués de tumeurs lumineales avec une forte instabilité génomique. Le pronostic de ces groupes est qualifié d'intermédiaire avec une survie à 10 ans de 50 à 60%. Les tumeurs du cluster 6, principalement lumineales A, sont caractérisées par une amplification du locus 8<sup>12</sup>. La fréquence de mutation du gène *PIK3CA* est la plus faible au sein de ce cluster. Le cluster 9, enrichi en tumeurs lumineales B, est caractérisé par un gain en 8q et une amplification en 20q. Dans ces tumeurs, principalement de haut grade, le gène *TP53* est fréquemment muté (≈65%). Enfin, les tumeurs du cluster 1 sont très majoritairement du groupe Luminal B. Leur génome montre une amplification du locus 17q23 et du bras long du chromosome 20q. Le gène *GATA3* est fréquemment muté (>25%).

Le cluster 5 est défini par une amplification du locus 17q12 et un mauvais pronostic avec une survie à 10 ans estimée à 45%. Il contient les tumeurs ERBB2+ et quelques tumeurs lumineales. Les tumeurs sont de haut grade avec une forte instabilité génétique. Le gène *TP53* est muté dans plus de 60% des cas.

Le cluster 10 rassemble la majorité des tumeurs triple-négatives du sous-type intrinsèque basal-like. Ces tumeurs sont de haut grade, peu différenciées et sont principalement retrouvées chez les femmes jeunes. Le taux de mutation de *TP53* dans ce cluster est le plus élevé, l'index mitotique est élevé, mais l'instabilité génomique est qualifiée d'intermédiaire. Le pronostic de survie à 5 ans est faible, mais devient intermédiaire à 10 ans. Ces tumeurs présentent une perte en 5q et des gains en 8q, 10 et 12. La délétion en 5q est associée à la modulation de l'expression de gènes situés à distance (activité *trans*) contrôlant le cycle cellulaire, la réparation de l'ADN et l'apoptose. Curtis *et al*, proposent que la perte du 5q joue un rôle dans l'instabilité génomique et la dérégulation du cycle cellulaire, expliquant la prolifération élevée observée dans ce groupe [11].

Enfin, le cluster 4 se distingue des autres par sa composition « mixte ». Il regroupe les tumeurs ER+ et ER-, dont 26% des tumeurs triple-négatives. Ce sous-groupe est associé à un pronostic favorable à 10 ans de l'ordre de 80% et est caractérisé par un profil génomique lat. Ce groupe présente également un infiltrat lymphocytaire important.

Table II Features of the integrative clusters

IntClust	Frequency (n, %)	Defining molecular features	Expression (n, %)	PAM50 (n, %)	Clinical features	Prognosis (5-year, 10-year DSS)	Genomic instability
1	139 (7%)	17q23 amplification	ERb : 123 (88.49%) PRb : 60 (43.17%) HER2b : 20 (14.39%)	Basal: 9 (6.47%) HER2: 21 (15.11%) LumA: 11 (7.91%) LumB: 90 (64.75%) Normal: 8 (5.76%) Basal: 2 (2.78%) HER2: 6 (8.33%) LumA: 25 (34.72%) LumB: 36 (50%) Normal: 3 (4.17%) Basal: 4 (1.39%) HER2: 9 (3.14%) LumA: 195 (67.94%) LumB: 43 (14.98%) Normal: 36 (12.54%) Basal: 64 (18.71%) HER2: 34 (9.94%) LumA: 106 (30.99%) LumB: 29 (8.48%) Normal: 109 (31.87%) Basal: 21 (11.05%) HER2: 108 (56.84%) LumA: 18 (9.47%) LumB: 33 (17.37%) Normal: 10 (5.26%) Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	High grade	Intermediate 0.80, 0.69	High
2	72 (4%)	11q13/14 amplification	ERb : 69 (95.83%) PRb : 51 (70.83%) HER2b : 3 (4.17%)	Basal: 2 (2.78%) HER2: 6 (8.33%) LumA: 25 (34.72%) LumB: 36 (50%) Normal: 3 (4.17%) Basal: 4 (1.39%) HER2: 9 (3.14%) LumA: 195 (67.94%) LumB: 43 (14.98%) Normal: 36 (12.54%) Basal: 64 (18.71%) HER2: 34 (9.94%) LumA: 106 (30.99%) LumB: 29 (8.48%) Normal: 109 (31.87%) Basal: 21 (11.05%) HER2: 108 (56.84%) LumA: 18 (9.47%) LumB: 33 (17.37%) Normal: 10 (5.26%) Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	No distinct clinical features	Poor 0.78, 0.51	High
3	290 (15%)	Paucity of copy number changes	ERb : 278 (95.86%) PRb : 211 (72.76%) HER2b : 1 (0.34%)	Basal: 4 (1.39%) HER2: 9 (3.14%) LumA: 195 (67.94%) LumB: 43 (14.98%) Normal: 36 (12.54%) Basal: 64 (18.71%) HER2: 34 (9.94%) LumA: 106 (30.99%) LumB: 29 (8.48%) Normal: 109 (31.87%) Basal: 21 (11.05%) HER2: 108 (56.84%) LumA: 18 (9.47%) LumB: 33 (17.37%) Normal: 10 (5.26%) Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Low grade Low LNP	Good 0.93, 0.88	Low
4	343 (17%)	CNA devoid	ERb : 238 (69.39%) PRb : 155 (45.19%) HER2b : 20 (5.83%)	Basal: 64 (18.71%) HER2: 34 (9.94%) LumA: 106 (30.99%) LumB: 29 (8.48%) Normal: 109 (31.87%) Basal: 21 (11.05%) HER2: 108 (56.84%) LumA: 18 (9.47%) LumB: 33 (17.37%) Normal: 10 (5.26%) Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Low grade	Good 0.89, 0.76	Low
5	190 (10%)	ERBB2 amplification	ERb : 79 (41.58%) PRb : 40 (21.05%) HER2b : 181 (95.26%)	Basal: 21 (11.05%) HER2: 108 (56.84%) LumA: 18 (9.47%) LumB: 33 (17.37%) Normal: 10 (5.26%) Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Younger age at diagnosis High grade High LNP	Poor 0.62, 0.45	Intermediate
6	85 (4%)	8p12 amplification	ERb : 85 (100%) PRb : 36 (45.88%) HER2b : 3 (3.53%)	Basal: 3 (3.53%) HER2: 10 (11.76%) LumA: 23 (27.06%) LumB: 43 (50.59%) Normal: 6 (7.06%) Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	No distinct clinical features	Intermediate 0.83, 0.59	High
7	190 (10%)	16p gain, 16q loss, 8q amplification	ERb : 187 (98.42%) PRb : 150 (78.95%) HER2b : 2 (1.05%)	Basal: 3 (1.59%) HER2: 9 (4.76%) LumA: 123 (65.08%) LumB: 41 (21.69%) Normal: 13 (6.88%) Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Older age at diagnosis Low grade	Good 0.94, 0.81	Intermediate
8	299 (15%)	1q gain, 16q loss	ERb : 297 (99.3%) PRb : 236 (78.93%) HER2b : 1 (0.33%)	Basal: 1 (0.33%) HER2: 9 (3.01%) LumA: 192 (64.21%) LumB: 89 (29.77%) Normal: 8 (2.68%) Basal: 20 (13.79%) HER2: 26 (17.93%) LumA: 24 (16.55%) LumB: 70 (48.28%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Older age at diagnosis Low grade	Good 0.88, 0.78	Intermediate
9	146 (7%)	8q gain, 20q amplification	ERb : 125 (85.62%) PRb : 79 (54.11%) HER2b : 10 (6.85%)	Basal: 8 (2.68%) HER2: 20 (13.79%) LumA: 26 (17.93%) LumB: 24 (16.55%) Normal: 5 (3.45%) Basal: 202 (89.38%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	High grade	Intermediate 0.78, 0.62	High
10	226 (11%)	5q loss, 8q gain, 10p gain, 12p gain	ERb : 25 (11.06%) PRb : 19 (8.41%) HER2b : 6 (2.65%)	Basal: 5 (3.45%) HER2: 8 (3.54%) LumA: 1 (0.44%) LumB: 14 (6.19%) Normal: 1 (0.44%)	Younger age at diagnosis High grade Large tumours	Poor 0.71, 0.68	Intermediate

IntClust, integrative cluster; DSS, disease-specific survival; LNP, lymph-node involvement.

**Table 1 : Description des clusters intégratifs issus de l'analyse des données génomiques et transcriptomiques de près de 2000 cancers du sein (Extrait de Dawson et al, 2013)[84]. Les 10 clusters intégratifs sont décrits selon les items suivants : fréquence dans la cohorte, altérations génomiques spécifiques, expression des récepteurs hormonaux RE (œstrogène) et RP (progestérone), expression de HER2, classification intrinsèque des tumeurs selon le test PAM50, caractéristiques cliniques (grade et envahissement ganglionnaire –LN pour lymph-node involvement), pronostic de survie à 5 ans et 10 ans (DSS ; disease-specific survival), instabilité génomique.**

### 1.3. Prédiposition génétique au cancer du sein

L'existence d'une composante héréditaire dans le cancer du sein est suissée depuis très longtemps comme en témoigne le Traité des tumeurs de Paul Broca de 1866 qui rapporte le cas d'une agrégation familiale de cancers du sein ainsi que la mise en évidence d'un risque tumoral mammaire augmenté de femmes dont la mère ou la sœur a été atteinte de cancer du sein [85]. Le modèle de prédiposition génétique au cancer a ensuite été proposé par Knudson [86;87]. Pour aboutir à une cellule tumorale, la théorie des deux événements de Knudson postule l'existence d'une première mutation constitutionnelle (héritée des parents ou de novo) sur un allèle, suivie d'une seconde mutation somatique sur le second allèle d'une même cellule. Ce modèle découvert dans le cas du rétinoblastome s'applique au syndrome de prédiposition héréditaire aux cancers par la mutation constitutionnelle d'un gène de prédiposition au cancer.

#### 1.3.1. Les gènes de prédiposition au cancer du sein

Des avancées majeures ont été réalisées dans l'étude des cancers du sein dans les années 1990 grâce à des études de liaison génétique qui étudient la coségrégation de marqueurs génétiques dans le but de localiser les gènes par la méthode des *lod score* [88]. Ces analyses 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 locus d'intérêt. Si deux locus sont proches sur le génome, l'haplotype se transmet comme un seul gène, en revanche si les locus sont éloignés ou situés sur deux chromosomes différents, on observera une ségrégation indépendante des locis.

Ces études de liaison ont conduit à la localisation en 17q21 du premier gène de prédiposition au cancer du sein *BRCA1* (*Breast Cancer 1*) en 1990 cloné en 1994 [89;90] et peu de temps après, à la découverte de *BRCA2* (*Breast Cancer 2*) dans la région 13q12-13 et de son clonage [91]. Ces gènes sont considérés comme gènes majeurs de prédiposition au cancer du sein et regroupent environ 20 à 30% des formes familiales de cancers du sein [92;93]. Les facteurs génétiques impliqués dans l'augmentation de risque de cancer du sein identifiés à ce jour sont classés en fonction de leur pénétrance [94].

On distingue :

(i) les gènes associés à une forte pénétrance : la fréquence dans la population générale d'une mutation des gènes *BRCA1* et *BRCA2* est de 0.2% et confère un risque cumulé de 40 à 85%.

On retrouve également dans cette catégorie les gènes *TP53* (gène de prédiposition au syndrome de Li-Fraumeni), *PTEN* (gène de prédiposition au syndrome de Cowden) et *STK11* (sérine/thréonine kinase

11, gène de prédisposition au syndrome de Peutz-Jeghers) dont la fréquence est inférieure à 0.01% et qui concernent moins de 1% des prédispositions au cancer du sein. Seuls ces gènes font l'objet d'un diagnostic moléculaire chez les patientes susceptibles de présenter une prédisposition génétique [92]. Le risque relatif de ces gènes, inférieur à celui des gènes *BRCA1* et *BRCA2*, est évalué entre 2 et 10.

(ii) les gènes associés à une pénétrance modérée : la fréquence de mutation de ces gènes dans la population générale est estimée entre 0.1 et 0.5%. On retrouve dans cette catégorie les gènes *ATM* (*ataxia telangiectasia mutated*), *BARD1* (*BRCA1-associated RING domain protein*) [95], *BRIP1* (*BRCA1 interacting protein C-terminal helicase 1*), *CHEK2* (*checkpoint kinase 2*) et le gène *PALB2* (*partner and localizer of BRCA2*). Le risque relatif associé à une mutation d'un de ces gènes est compris entre 2 et 3 sauf pour le gène *PALB2* qui est estimé à 6 [96]. Récemment, le gène *RAD51C* a été reporté comme gène de prédisposition au cancer du sein et de l'ovaire. Thomson *et al.*, suggèrent que la mutation de ce gène est retrouvée dans moins de 1% des familles présentant un syndrome sein ovaire, mais aucune mutation n'a été identifiée dans les familles présentant uniquement une agrégation de cas de cancers du sein [97-99].

(iii) les gènes associés à une faible pénétrance : risque relatif inférieur à deux (risque relatif compris entre 1.25 à 1.65) et à une fréquence modérée à élevée (5 à 50%) ont été découverts par l'intermédiaire de *Genome-wide association study* (GWAS) utilisant de nouvelles techniques de génétique à grande échelle, de type *SNP* [100-104]. Les GWAS combinent les résultats de génétique de populations (population d'étude versus population témoin) dans le but d'identifier les gènes associés à une maladie, ici le cancer du sein. Les auteurs de ces études identifient entre autre les gènes *FGFR2* (*fibroblast growth factor receptor 2*, rs2981582), *RAD51L1* (*RAD51 Homolog B*, rs999737) *TOX3* (*TOX high mobility group box family member 3*, rs3803662), *MAP3K1* (rs889312), *LSP1* (*lymphocyte-specific protein 1*, rs3817198), *CASP8* (caspase 8, rs1045485) et *TBX3* (*T-box3*, rs1292011). Ces anomalies sont retrouvées dans 3.9% des cas de cancers familiaux [105] mais ne peuvent être considérées comme responsables de formes familiales de cancers du sein en raison d'une pénétrance trop faible.

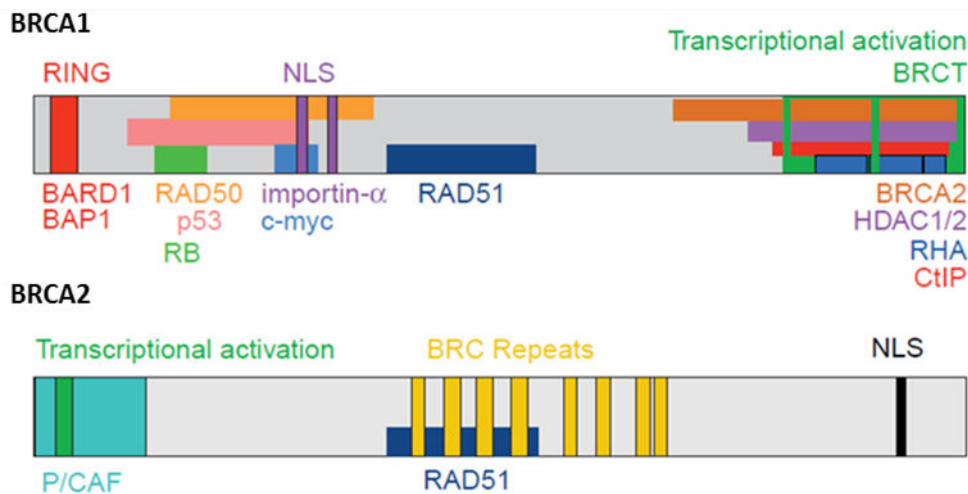
### 1.3.2. Les gènes majeurs de rédisposition *BRCA1* et *BRCA2*

Les gènes *BRCA1* et *BRCA2* localisés respectivement en 17q21 et 13q13, ne présentent aucune identité de séquence. *BRCA1* s'étend sur 81 kb (kilobases) et est constitué de 24 exons (dont 22 sont codants). *BRCA2* s'étend sur 84 kb et 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*.

Le gène *BRCA1* code pour une protéine de 1863 AA (acides aminés) (Figure 3). La partie carboxy-terminale de *BRCA1* possède un motif polypeptidique nommé domaine BRCT (*BRCA1 C-Terminus*) réitéré en tandem, présent dans d'autres protéines de réparation de l'ADN [106]. Le tandem de deux domaines BRCT semble impliqué dans les interactions protéine-protéine telles que les protéines *BRCA2*, *RHA (RNA Helicase A)*, *CtIP (CtBp-interacting protein)* et *HDAC 1 et 2* [107;108]. Un domaine en doigts de zinc, appelé RING Finger, impliqué dans l'ubiquitination est présent en aminoterminal [109]. Ce motif riche en cystéine et histidine est caractéristique des protéines présentant une activité ubiquitine ligase ou E3 ligase, et permet de lier deux atomes de zinc [110]. L'ubiquitination est un processus multi-étape (activation, conjugaison, transfert au substrat) faisant intervenir successivement trois grandes classes d'enzymes appelées E1-, E2- et E3-ligases, permettant l'adressage d'un substrat modifié lié de façon covalente à l'ubiquitine, petite protéine de 76 AA, vers le protéasome pour sa dégradation. Le domaine RING est le siège de l'hétérodimérisation entre *BRCA1* et *BARD1*, qui possède également un domaine RING et deux domaines BRCT. Le recrutement de la protéine *BARD1* permet l'interaction du domaine RING *finger* *BRCA1* avec l'ADN. Le domaine RING fixe également *BAP1 (BRCA1-associated protein 1)*, gène impliqué dans la rédisposition au cancer codant pour une enzyme présentant une activité d'ubiquitine-hydrolase [108]. Une région située en 5' de l'exon 11 comportant deux signaux de localisation nucléaire ou NLS (*Nuclear Localisation Signal*) semblerait également impliquée dans les interactions protéine-protéine [111].

Les domaines BRCT et RING sont conservés au cours de l'évolution, et l'observation de mutations intervenant dans ces domaines associées à une rédisposition au cancer du sein démontre leur importance dans la fonction de *BRCA1* [112].

Le gène *BRCA2* code pour une protéine de 3418 AA dont la fonction et la structure seront détaillées dans le paragraphe 1.4.1.. Brièvement, *BRCA2* possède un domaine de transactivation aminoterminal, 8 domaines BRC (*BRCA2-repeats*) situés dans la partie centrale codés par l'exon 11 et un signal de localisation nucléaire carboxy-terminal. Les domaines BRC sont très conservés et sont capables de se lier à la protéine *RAD51A* [113], protéine essentielle dans la réparation des cassures double brin (Figure 3) [108;114].



**Figure 3: Structure et protéines partenaires des protéines BRCA1 et BRCA2** (adapté de Welch et al 2000) [108].

Les mutations des gènes *BRCA1* et *BRCA2* sont associées à une `rédis`osition au cancer du sein et de l'ovaire. Le risque cumulé à 70 ans `our le cancer du sein est de 65% `our *BRCA1* et de 45% `our *BRCA2* contre 10% dans la `o`ulation générale [115]. Pour le cancer de l'ovaire, le risque cumulé à 70 ans est de 39% `our *BRCA1* et de 11% `our *BRCA2* [115]. Le risque relatif (RR) de cancer du sein et de l'ovaire liés aux gènes *BRCA1* et *BRCA2* évolue avec l'âge. En effet, le RR de cancer du sein diminue avec l'âge `our les `orteurs de mutation du gène *BRCA1*, et le RR de cancer de l'ovaire augmente avec l'âge `our les `orteurs de mutation du gène *BRCA2*. Les cancers du sein masculins sont `lus souvent associés à une `rédis`osition *BRCA2* et le risque cumulé à 70 ans chez l'homme est estimé à 6% [116].

Au-delà de la `rédis`osition au cancer du sein et de l'ovaire, des études menées `ar le "Breast Cancer Linkage Consortium" (BCLC) chez les `orteurs de mutation des gènes *BRCA1/2* ont ra`orté une association à d'autres cancers. Ainsi, dans un contexte *BRCA1*, le cancer de la `rostate chez l'homme de moins de 65 ans, le cancer du côlon, du `ancréas et du col de l'utérus sont associés à un risque relatif de 1.8, 2.0, 2.3, et 3.7 respectivement [117]. Dans un contexte *BRCA2*, le risque relatif de dévelo`er un cancer de la `rostate est `orté à 4.6 et celui du `ancréas à 3.5. L'augmentation de risque de mélanome et de cancer de l'estomac est également observée [118;119].

Certaines études su`ortent l'idée que la `osition de la mutation `ourrait modifier le risque de cancer. Pour le gène *BRCA1*, les mutations dans la région centrale (exon 11), seraient associées à une diminution du risque de cancer du sein et une augmentation du risque de cancer l'ovaire [120]. Si les mutations se situent dans la région OCCR (Ovarian Cancer Cluster Region) de l'exon 11 du gène

*BRCA2*, le risque de cancer du sein serait plus faible, et celui de l'ovaire serait plus élevé [121]. Cependant ces données n'ont pas été réellement confirmées.

Cette rédisposition suit le modèle de Knudson [86] par la mutation constitutionnelle d'un allèle et l'inactivation somatique de l'allèle sauvage conduisant ainsi à l'inactivation bi-allélique de ces 2 gènes dans les cellules cancéreuses de tumeurs survenant dans ce cadre [122]. Les produits des gènes *BRCA1* et *BRCA2* sont dans les deux cas impliqués dans la réparation des dommages à l'ADN, les faisant considérer comme des gènes suppresseurs de tumeur de type *caretaker*. Cette famille de gènes informe la cellule des lésions de l'ADN et peut participer aux processus de réparation [123;124].

L'analyse des gènes *BRCA1* et *BRCA2* dans les cancers sporadiques du sein et de l'ovaire a montré une perte d'hétérozygotie au locus *BRCA1* dans 50% des cas, contre 30% au locus *BRCA2* [125;126]. La diminution d'expression des protéines BRCA1 et BRCA2 a été observée dans la majorité des tumeurs du sein et de l'ovaire étudiées [127;128]. En revanche, l'inactivation somatique des gènes *BRCA1* et *BRCA2* par mutation ponctuelle et délétion est rare [129].

L'hyperméthylation somatique du promoteur du gène *BRCA1* conduisant à une réduction d'expression a été observée dans 15% des cancers sporadiques, mais semble absente pour le gène *BRCA2* [130-132].

### 1.3.3. Caractéristiques des tumeurs du sein liées à une mutation constitutionnelle de *BRCA2*

#### 1.3.3.1. Clinique : morphologie, histologie, grade

Différentes études s'accordent à dire que les Carcinomes Canalaires Infiltrants (CCI) représentent la majorité des tumeurs BRCA2, soit environ 76% selon l'étude du Breast Cancer Linkage Consortium (BCLC) réalisée sur 78 tumeurs BRCA2 [133].

Les cancers du sein liés à une mutation constitutionnelle de *BRCA2* sont majoritairement des tumeurs de haut grade, grade II ou III (Table 2). De nombreuses études s'accordent à dire que la différenciation tubulaire des tumeurs BRCA2 est faible (score de différenciation de 3). Selon l'équipe de Lakhani et l'étude du Breast Cancer Linkage Consortium, le haut grade SBR serait uniquement dû à la faible différenciation glandulaire [133;134]. Le léomorphisme nucléaire est modéré et le score mitotique élevé, significativement supérieur au groupe contrôle [135].

Critères	Contrôles sporadiques (%)	BRCA2 (%)	Références
Grade III	39	60	[135]
RE-négative	29	22	[135]
HER2 négative	85	94	[135]
Amplification de <i>MYC</i>	23-31	62.5	[136]
Mutations de <i>TP53</i>	35	63	[136]

**Table 2 : Distribution des caractéristiques différenciant les tumeurs du sein liées à BRCA2 des tumeurs du sein sporadiques** (Adapté de Bane et al., [135] et Turner et al., [136]).

### 1.3.3.2. Classification hétéotypique

Les tumeurs BRCA2 présentent majoritairement une positivité aux récepteurs hormonaux ainsi qu'une absence de surexpression des protéines ERBB2 et des cytokératines basales CK5/6 [135;137;138]. Ces critères classent les cancers du sein liés à BRCA2 dans le hétéotype Luminal malgré leur haut grade histologique [135;139]. Elles sont le plus souvent mutées pour TP53 comparées aux tumeurs sporadiques et de nombreuses études ont mis en évidence la surexpression de protéines impliquées dans la carcinogénèse comme la Cycline D1 [140] et MYC [136;141].

### 1.3.3.3. Profils génomique et transcriptomique des tumeurs BRCA2

L'instabilité génétique qui caractérise ces 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 3 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 [139;142-145]. 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), présente une sensibilité de 89 % et une spécificité de 84 % [144].

Les études transcriptomiques quant à elles identifient les tumeurs BRCA2 comme appartenant essentiellement au groupe Luminal B de la classification intrinsèque d'expression de Sorlie [68]. En



effet, Waddell et son équipe ont observé que 40% des tumeurs BRCA2 étudiées en transcritome (n=30) sont lumineales B et 33% sont lumineales 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 [146].

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 permuation randomisé révèle seulement 4% de mauvaise classification. Cependant cette signature n'a pas été validée sur un set de tumeurs indépendant [147]. Ces deux signatures des tumeurs BRCA2 proposées par Hedenfalk et Waddell n'ont aucun gène en commun.

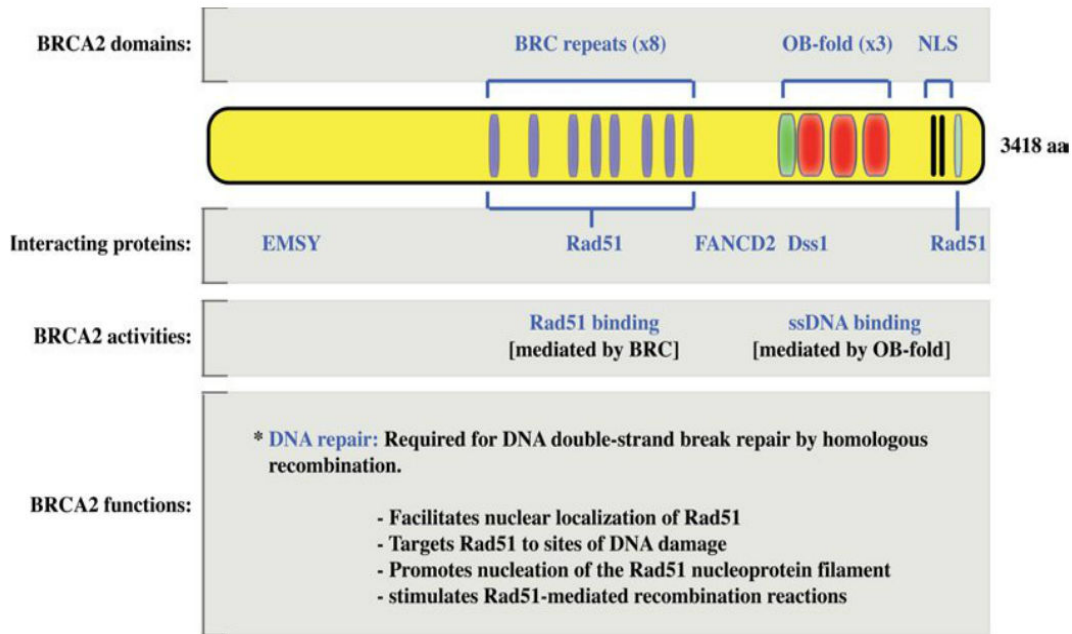
## 1.4. Fonctions cellulaires de la protéine BRCA2

### 1.4.1. La protéine BRCA2 : structure et transcription

La protéine BRCA2 est un des plus grands polypeptides du protéome humain (3418 AA). Cette protéine est formée d'un motif BRC, répété huit fois sur 1000 AA, chaque domaine BRC étant composé d'environ 70 AA. Les domaines BRC ont été démontré capables de se lier à la protéine RAD51 avec une affinité différente selon des études de double-hybride réalisées chez la levure. Ainsi, les domaines BRC5 et BRC6 ont une plus faible affinité que les autres [114;148].

Des analyses de structure ont révélé 5 domaines distincts dans la partie carboxy-terminale de la protéine. Le premier domaine est un domaine hélicoïdal de 190 AA identifié par Yang qui a la particularité de se lier au polypeptide DSS1 (*Deleted In Split Hand/Split Foot Protein 1*) et qui augmenterait la solubilité de la protéine [149]. Ce domaine est suivi de trois structures répétées d'environ 110 AA (OB1, OB2 et OB3) qui présentent une homologie de structure avec les domaines OB-fold (*oligonucleotide/oligosaccharide-binding*) présents chez la plupart des eucaryotes et procaryotes. Ces domaines OB-fold sont retrouvés dans les protéines de liaison à l'ADN simple brin comme les SSB (*ssDNA-binding protein*) ou RPA (*replication protein A*). Contrairement aux domaines OB1 et OB3, le domaine OB2 contient une séquence de 130 AA appelée « tower domain » en rapport avec sa structure tridimensionnelle en forme de tour, se liant à l'ADN double brin [149] (Figure 4).

En plus des domaines BRC, un second site de liaison à la protéine RAD51, est présent en C-terminal [150;151]. Ce site contient la Sérine<sup>3291</sup> qui, selon son état de phosphorylation par la Cdk (*cyclin-dependent kinase*), module la liaison de RAD51 à ce dernier [152]. Enfin, deux signaux (NLS) situés dans la région C-terminale de la protéine BRCA2, les acides aminés 3266 à 3270 et 3311 à 3316, sont essentiels à sa localisation nucléaire.



**Figure 4 : Protéine BRCA2 : domaines, protéines partenaires, activités et fonctions** (Extrait de Boulton et al., 2006) [123].

Le gène *BRCA2* possède une séquence codante de 10.2 kb et contient 27 exons, dont l'exon 11 de 4.9 kb. Chez l'homme, le transcrit *BRCA2* est exprimé préférentiellement dans le thymus et les testicules mais également dans de nombreux tissus tels que le sein et l'ovaire [153].

La vaste majorité des mutations délétères introduisent un codon stop prématuré (Base BIC) ce qui entraîne la dégradation du transcrit par le système Nonsense-Mediated mRNA Decay (NMD) ou la synthèse d'une protéine non fonctionnelle [154].

Dans le gène *BRCA2*, peu de transcrits alternatifs sont connus. Actuellement les transcrits alternatifs physiologiques connus ont perdu un ou deux exons (nommés « delta ») et sont : delta-exon 3, delta-exons 6-7, delta-exon 12, et delta-exons 17-18 [155].

L'expression du transcrit delta-exon 12-*BRCA2* serait plus élevée dans les cellules issues de tumeurs mammaires [156].

#### 1.4.2. Rôles de la protéine BRCA2

La protéine BRCA2 semble impliquée dans la détection et/ou la réparation des lésions de l'ADN, en particulier dans la recombinaison homologue. Ces rôles la font considérer comme « *caretaker* » car elle a la charge de contrôler et de maintenir l'intégrité du génome [157]. La perte de ces fonctions conduit à l'instabilité génétique comme en témoigne le caryotype de cellules déficientes en BRCA2 par la présence de nombreuses translocations, duplications et fusions aberrantes entre chromosomes non-homologues. Les embryons de souris BRCA2<sup>-/-</sup> meurent à un stade précoce et sont hypersensibles aux radiations  $\gamma$  induisant des dommages à l'ADN [158].

La formation de foci de réparation par BRCA2 sous l'action d'agents alkylants laisse à penser que la protéine aurait un rôle direct dans la réparation de l'ADN double brin [148]. Trois voies de réparation de ces cassures sont connues actuellement : la recombinaison homologue (HR, *Homologous Recombination*), l'hybridation simple brin (SSA, *Single Strand Annealing*) et la jonction de brins non homologues (NHEJ, *Non Homologous End Joining*), ces deux dernières introduisant des erreurs car peu fidèles. BRCA2 serait impliquée uniquement dans la recombinaison homologue [159].

L'expression de BRCA2 est dépendante du cycle cellulaire avec un pic d'expression durant les phases S et G<sub>2</sub>. De faibles niveaux d'expression sont détectés durant les phases G<sub>0</sub>, G<sub>1</sub> et M. Ces résultats impliquent que la protéine BRCA2 est utile à la croissance et peut avoir un rôle protecteur dans la prolifération cellulaire [160].

Le partenaire privilégié de BRCA2 est la protéine RAD51. Cette dernière est connue pour intervenir dans la réparation des cassures double brin de l'ADN et dans la recombinaison homologue associée à la réparation. BRCA2 interagit directement avec RAD51 par l'intermédiaire de ces différents sites de fixation, les domaines BRC et le domaine C-terminal [114;123;148]. BRCA2 assurerait la localisation de RAD51 aux sites de cassure de l'ADN.

Un second partenaire de BRCA2 est la protéine DSS1 identifiée lors d'une expérience de double hybride chez la levure par Marston et al. [161]. La délétion de DSS1 dans les cellules humaines inhibe fortement la formation des foci de RAD51 au niveau des cassures double brin [162]. DSS1 est donc nécessaire à la machinerie de réparation des cassures double brin via la recombinaison homologue régie par RAD51. Ainsi, BRCA2 est indispensable au recrutement de RAD51 aux sites de cassure et intervient dans la liaison ADN simple brin-ADN double brin. BRCA2 est directement impliquée dans la réparation par recombinaison homologue en catalysant la formation de filaments de RAD51 sur les brins d'ADN au niveau des points de cassure [163]. BRCA2 participe également au changement de brin de RAD51 et à la formation de la D-loop ou « bulle de rélication » (« *Displacement-loop* »). L'ADN simple brin contenu dans le nucléo-filament forme des paires de bases Watson-Crick avec le brin complémentaire du duplexe « cible » [113] (Figure 5).

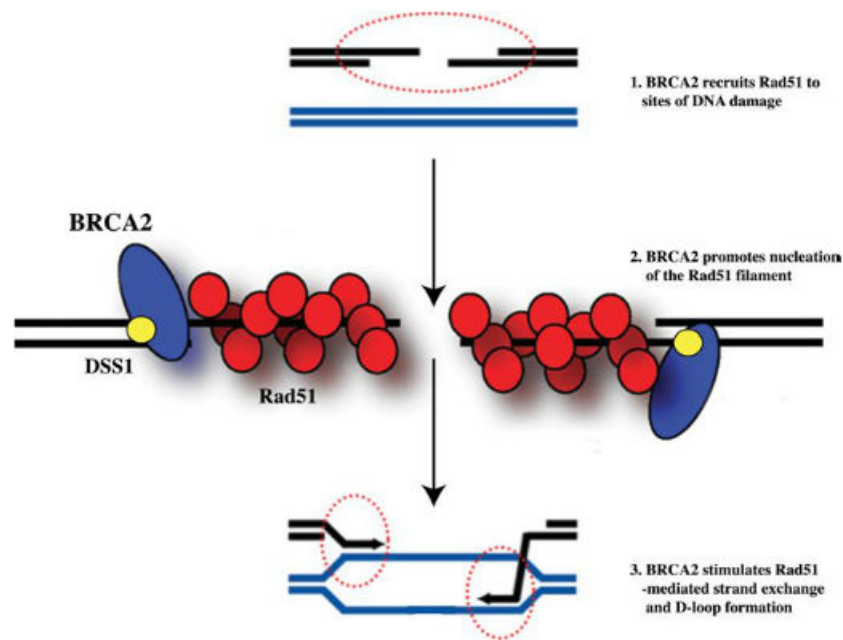
D'après une étude de double-hybride chez la levure et de co-immunoprécipitation la protéine FANCD2 (Fanconi anemia, complémententation groupe D2), appartenant au complexe de réparation de l'ADN de l'Anémie de Fanconi (FA), se lie directement à BRCA2 [164]. De plus, l'inactivation bi-allélique de BRCA2 a été montrée comme étant responsable d'un des groupes de complémententation de l'anémie de Fanconi (groupe D1). Cette observation lie de nouveau la protéine BRCA2 à la recombinaison homologue orchestrée par le système FA. S'ajoute à cela diverses études démontrant que la perte de fonction du système FA dans les lignées cellulaires entraîne la déficience en recombinaison homologue des cassures double brin [165;166].

Enfin, deux protéines ont à leur tour été identifiées comme interagissant avec BRCA2, la protéine EMSY et la protéine PALB2. EMSY s'associerait à BRCA2 au niveau de la région de l'exon 3 de BRCA2 qui code pour le domaine de transactivation pour exercer une régulation négative sur l'activité de BRCA2. EMSY serait recrutée au niveau des sites de cassures, mais son rôle dans la réparation reste inconnu [167]. PALB2 contrôle la localisation de BRCA2 et maintient sa stabilité au niveau de la chromatine ; elle semble 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 [168].

La protéine BRCA2 serait également impliquée dans la régulation de la transcription. L'exon 3 du gène présente une homologie de séquence avec le domaine activateur du facteur de transcription c-Jun liant JNK (*Jun N-terminal kinase*) [169]. De plus, la partie N-terminale de BRCA2 interagit avec le complexe P/CAF ayant une activité histone acétyltransférase [170].

Daniels *et al*, [171] et Jonsdottir *et al*, [172] ont étudié le lien entre la cytokinèse et BRCA2. L'inactivation de la protéine BRCA2 dans des cultures primaires de fibroblastes humains et murins entraînerait un dysfonctionnement de la cytokinèse. Les défauts de séparation des cellules filles seraient à l'origine d'une mauvaise ségrégation des chromosomes et entraînerait l'instabilité génomique, l'aneuploïdie et l'amplification des centrosomes [173]. Ces résultats montrent que BRCA2 joue un rôle clé dans la croissance cellulaire et la prolifération.

Enfin, l'interaction de BRCA2 avec la protéine MAGE-D1 (*Melanoma Antigen Family D1*) serait à l'origine d'un arrêt de la prolifération cellulaire [174]. Une étude datant de 2002 indique que MAGE-D1 serait un régulateur de l'apoptose, de la transcription et de la progression du cycle cellulaire [175].



**Figure 5 : Modèle de la fonction de BRCA2 dans la réparation des cassures double-brin par recombinaison homologue** (Extrait de Boulton *et al.*, 2006) [123]. (1) BRCA2 est nécessaire au recrutement de Rad51 aux cassures double brin. (2) BRCA2 intervient dans la liaison ADN simple brin-ADN double brin et facilite la formation des filaments de Rad51. (3) Une fois les filaments de nucléoprotéine Rad51 formés, BRCA2 stimule le changement de brin et la formation de la D-loop.

### 1.5. Développement d'une thérapie ciblée : le traitement antiPARP

L'enzyme PARP1 (poly(ADP-ribose) polymérase 1) est une ADN polymérase qui contribue à maintenir l'intégrité de l'ADN en éliminant les cassures simple brin ou les bases endommagées, irréversibles au fonctionnement des cellules par un système de réparation de l'ADN, le système BER (Base Excision Repair) [176]. De nouvelles molécules en essai clinique exploitent les défauts de recombinaison homologue en thérapie ciblée : les inhibiteurs de PARP (antiPARP). Des défauts de recombinaison homologue sont induits par la perte de BRCA2 au niveau fonctionnel [173;177;178]. La cellule tumorale semble utiliser alors comme système de secours de réparation de l'ADN, le système BER. En inhibant ce dernier (par inhibition de PARP1), les cellules tumorales qui présentent une inactivation bi-allélique de *BRCA2* entrent en apoptose comme en témoignent les expériences de l'équipe de Bryant *et al.*, conduites sur des cellules MCF7 traitées par des inhibiteurs de PARP1 et un siRNA dirigé contre *BRCA2* [178]. Dans les cellules déficientes en BRCA2, le défaut de réparation par recombinaison homologue associée à l'inhibition de PARP1 induit de façon sélective et massive la

mort cellulaire selon un mécanisme de synergie létale a`elée aussi létalité synthétique de l'anglais *synthetic lethality* [176;178].

Une étude de phase 1 conduite par Fong *et al.*, a testé la tolérance de l'olaparib, un inhibiteur de PARP1 sur 60 personnes atteintes de divers cancers [179]. Vingt-et-une patientes étaient atteintes d'un cancer de l'ovaire (15 porteuses d'une mutation germinale de *BRCA1*, 1 de *BRCA2*) et 9 d'un cancer du sein (3 porteuses d'une mutation germinale de *BRCA2*). Les résultats ont montré une efficacité antitumorale chez 63% des patients porteurs de mutations *BRCA1* ou *BRCA2*. En 2010, une étude de phase 2 réalisée sur 57 patientes atteintes d'un cancer de l'ovaire récidivant, porteuses de mutation *BRCA1/2* confirme l'activité de l'olaparib sur les tumeurs liées à *BRCA1/2* avec une réponse partielle sous olaparib observée chez 33% des patientes [180]. Actuellement, le programme SOLO (*Study of Olaparib in Ovarian cancer*) entreprend une étude de phase 3 de l'olaparib dans les cancers de l'ovaire. Une publication de 2008 montre que la résistance acquise aux antiPARP dans les cellules tumorales déficientes en *BRCA2* peut être liée à la délétion d'une mutation tronquante de *BRCA2* qui restaure la phase. L'association de révertants lie l'efficacité des antiPARP à *BRCA2* [181].

Une publication introduit le concept de hénotypie BRCAness qui regroupe les tumeurs survenant chez des patients indemnes de mutation constitutionnelle de *BRCA1* ou *BRCA2*, mais présentant des caractéristiques hénotypiques, morphologiques et biologiques communes avec celles des patientes porteuses d'une mutation *BRCA1/2* [136]. La capacité à identifier ce groupe de tumeur BRCAness pourrait avoir un retentissement sur leur prise en charge, comme l'utilisation de la thérapie ciblée antiPARP. La recherche de caractéristiques tumorales histopathologiques ou moléculaires communes aux tumeurs présentant une altération des protéines *BRCA1/2* au sein des tumeurs est donc de haute importance pour élargir les possibilités thérapeutiques des antiPARP.

# OBJECTIFS

---

## 2. OBJECTIFS DE L'ETUDE

---

Les tumeurs BRCA2 constituent un groupe de tumeurs hétérogène. Aucun consensus concernant le type histologique et le profil d'expression de biomarqueurs des tumeurs BRCA2 n'est établi [135]. Plusieurs études récemment décrites ont montré que les profils génomiques des tumeurs BRCA2 présentent des lésions récurrentes de matériel génétique au niveau des bras longs des chromosomes 13 et 14, mais aucun profil génomique réellement spécifique n'a été jusqu'à présent caractérisé [139;142-145]. Cependant, la signification de ces remaniements est encore incertaine. Les études concernant le transcriptome de ces tumeurs héréditaires s'accordent pour classer les tumeurs BRCA2 dans le sous-type Luminal B en majorité [68;82].

L'indication d'une recherche de mutation constitutionnelle au niveau du gène *BRCA2* repose essentiellement sur des critères généalogiques de récurrence familiale de cancers du sein et de l'ovaire ((i) au moins 3 cas de cancers du sein chez des femmes apparentées au premier ou au second degré, ou (ii) au moins un cas de cancer du sein et un cas de cancer de l'ovaire, ou 2 cas de cancers du sein dont au moins un avant 36 ans, chez des femmes apparentées au premier degré (ou deuxième degré via un homme)). La fréquence de mise en évidence d'une mutation selon ces critères reste faible [182]. Selon le rapport de l'Inca de 2011, 4.9% des analyses du gène *BRCA2* ont conduit à l'identification d'une mutation [183]. En effet, le recrutement de patients sur des critères de récurrence familiale exclut les patients sans histoire familiale évocatrice. De plus, en raison de l'incidence des cancers du sein (évaluée à 99.7/100 000), il est possible que certaines agrégations familiales correspondent à des associations fortuites.

L'hétérogénéité des cancers du sein liés à *BRCA2* rend nécessaire l'établissement de profils génomiques et transcriptomiques spécifiques de tumeurs du sein survenues chez des patients mutés pour *BRCA2*, ceci dans une perspective diagnostique mais aussi thérapeutique en vue de l'utilisation d'une thérapie ciblée telle que le traitement antiPARP.

La définition de caractéristiques tumorales communes aux tumeurs du sein survenant dans un contexte de redispersion à *BRCA2* a deux objectifs :

(i) Etablir une signature moléculaire permettant de redire le statut mutationnel *BRCA2* sur des caractéristiques tumorales rendant éligible au diagnostic génétique les patients sans histoire familiale,



(ii) Explorer la signification des remaniements récurrents des tumeurs BRCA2 pouvant avoir un lien avec la tumorigénèse mammaire liée à une mutation constitutionnelle *BRCA2*.

Ce travail présente l'analyse conjointe des données transcriptomiques et génomiques d'une cohorte de tumeurs BRCA2 par mesure d'expression, CGH-array et mesure SNP. Les résultats de cette analyse ont ensuite conduit à l'exploration par séquençage nouvelle génération d'une portion de la cohorte.

Ce travail de thèse s'inscrit dans le cadre du programme CIT (Carte d'Identité des tumeurs) de la Ligue Contre Le Cancer et permet d'en utiliser les ressources.

# RESULTATS

---

## 3. RESULTATS

---

### 3.1. Caractérisation des échantillons de l'étude

L'étude inclut 33 échantillons de tumeurs du sein provenant du Centre de Ressources Biologiques de l'Institut Bergonié à l'exception d'un cas qui provient de l'Hôpital Val d'Aurelle à Montpellier et de 3 cas de l'Hôpital Dreyfus à Limoges (9 tumeurs BRCA2 et 24 tumeurs BRCA1 constituent le set d'entraînement). Nous avons par la suite collecté un nouveau set de données de tumeurs BRCA2 du sein en collaboration avec l'Institut Curie (IC) que nous avons enrichi avec d'autres échantillons BRCA1 provenant du Centre de Ressources Biologiques de l'Institut Bergonié (IB). Nous ajouterons ce nouveau set de données, set de validation (19 tumeurs BRCA2 et 12 tumeurs BRCA1).

#### 3.1.1. Criblage *BRCA1/BRCA2*

Le statut mutationnel des tumeurs du sein sélectionnées concernant les gènes de réduplication majeurs *BRCA1* et *BRCA2* a été caractérisé. Pour la majorité des échantillons, le criblage des gènes *BRCA1* et *BRCA2* avait déjà été réalisé lors de l'enquête oncogénétique entreprise dans la famille. Nous avons complété l'étude de ces gènes pour 13 patients. La recherche de mutations du gène *BRCA1* consiste en la recherche de réarrangements de grande taille qui correspondent à des délétions ou des duplications d'un segment génomique de taille variable par la technique de QMPSF (*Quantitative multiplex PCR on short fragments*) et la recherche de mutations ponctuelles par la technique EMMA (*Enhanced Mismatch Mutation Analysis*) [184-186]. La recherche de mutations du gène *BRCA2* dans le cadre de notre étude consiste uniquement en la recherche de mutations ponctuelles par la technique EMMA.

Les mutations des gènes de réduplication majeurs caractérisés sur l'ADN constitutionnel ont été ensuite recherchées sur le prélèvement tumoral correspondant dans un souci de vérification.

Ainsi le criblage des gènes *BRCA1* et *BRCA2* a permis de construire notre série d'étude BRCA2 avec 9 échantillons tumoraux représentant une mutation du gène *BRCA2* (Table 3).

Identification	Exon	Formule nucléotidique	Formule protéique
43	20	c.8537_8538delAG	p.Glu2846GlyfsX22
52	11	c.1929delG	p.Arg645GlyfsX15
86	11	c.5576_5579delTTAA	p.Ile1859LysfsX3
106	14	c.7069_7070delCT	p.Leu2357ValfsX2
133	19	c.8364G>A	p.Trp2788X
144	21	c.8639_8640delCA	p.Thr2880AsnfsX26
146	22	c.8904delC	p.Val2969CysfsX7
148	2	c.2T>G	p.Met1?
149	11	c.4889C>G	p.Ser1630X

**Table 3: Description des mutations constitutionnelles du gène *BRCA2* des 9 échantillons *BRCA2* de l'Institut Bergonié.** Les tumeurs 52, 86, 106, 133, 144, 146 et 148 constituent les tumeurs *BRCA2* du set d'entraînement.

### 3.1.2. Population BRCAX

Les tumeurs du groupe BRCAX sont des tumeurs du sein survenues dans un contexte héréditaire pour lesquelles aucune mutation des gènes de récidivation majeurs *BRCA1* et *BRCA2* n'a été caractérisée, soit par insuffisance de la technique utilisée soit en raison de la mutation d'autres gènes de récidivation non explorés. Notre étude dénombre 36 échantillons BRCAX. La sélection des patients BRCAX a été réalisée selon les critères de récurrence familiale suivants :

- \* Au moins 3 cas de cancers du sein chez des femmes apparentées au premier ou au second degré
- \* Au moins deux cas de cancer du sein dont un avant 42 ans ou un cas de cancer du sein et un cas de cancer de l'ovaire chez des femmes apparentées au premier degré (ou deuxième degré via un homme).

### 3.1.3. Caractéristiques anatomo-pathologiques et cliniques

La relecture des lames histologiques correspondant aux prélèvements tumoraux a été réalisée. Les dossiers cliniques ont été étudiés. Le pourcentage de cellules tumorales a été contrôlé par deux coupes en lame blanche sur le dessus et le dessous du prélèvement. Tous les échantillons tumoraux ont un pourcentage de cellules tumorales >50%. Les informations recueillies sur les échantillons tumoraux *BRCA2* et BRCAX sont présentées dans la Table 4.

Identification	Statut BRCA	Sexe	Age à la chirurgie	Taille de la tumeur (mm)	% CT	Type de Carcinome	Grade Histologique	IHC RE	IHC RP	IHC ERBB2
3	BRCAX	F	36	18	95	CCI	3	-	+	+++
8	BRCAX	F	51	18	90	CCI	3	-	-	-
9	BRCAX	F	51	25	95	CCI	3	++	++	-
11	BRCAX	F	56	40	78	CCI	2	++	+++	-
14	BRCAX	F	45	12	90	CCI	2	nd	+++	-
15	BRCAX	F	42	15	95	CCI	3	-	-	-
16	BRCAX	F	50	27	95	CCI	3	+++	+++	-
17	BRCAX	F	76	3	95	CCI	1	+++	+	-
22	BRCAX	F	64	18	90	CCI	2	+++	+	-
24	BRCAX	F	35	12	70	CCI	1	++	-	-
25	BRCAX	F	37	12	92	CCI	2	++	+	-
30	BRCAX	F	51	nd	95	CCI	3	+++	-	nd
33	BRCAX	F	42	35	73	CCI	1	++	-	-
37	BRCAX	F	45	20	92	CCI	2	+++	+++	-
38	BRCAX	F	64	13	90	CCI	3	+++	-	-
40	BRCAX	F	41	12	95	CCI	2	+	++	-
41	BRCAX	F	38	21	92	CCI	3	++	+	-
43	BRCA2	F	38	12	90	CCI	2	++	-	-
46	BRCAX	F	60	38	90	CAI	2	-	-	-
48	BRCAX	F	54	20	90	CCI	1	++	++	-
49	BRCAX	F	49	35	66	CCI	2	+++	++	-
52	BRCA2	F	35	17	92	CCI	3	++	-	-
65	BRCAX	F	46	37	95	CCI	3	-	-	-
66	BRCAX	F	73	12	90	CCI	2	+++	+++	-
71	BRCAX	F	43	21	73	CCI	2	++	+++	+++
75	BRCAX	F	58	14	80	CCI	2	-	-	+++
79	BRCAX	F	42	11	90	CCI	3	++	++	-
81	BRCAX	F	46	28	80	CCI et CLI	2	+	++	-
82	BRCAX	F	50	9	85	CCI	1	++	++	-
83	BRCAX	F	50	18	50	CLI	2	++	-	-
84	BRCAX	F	47	27	92	CCI	3	++	+	+
85	BRCAX	F	64	15	90	CCI	1	-	+++	-
86	BRCA2	F	46	16	90	CCI	3	+++	+	-
89	BRCAX	F	30	30	82	CCI	nd	++	+++	-
93	BRCAX	F	44	18	85	CCI	2	+++	+++	-
96	BRCAX	F	41	25	85	CCI	3	-	-	+++
99	BRCAX	M	63	21	90	CCI	1	+++	++	-
106	BRCA2	F	57	22	85	CCI	3	+++	+	-
107	BRCAX	F	69	40	80	CMuc	3	+++	+	-
111	BRCAX	F	73	15	80	CCI	1	+++	+	-
133	BRCA2	F	40	15	75	CCI	2	+	+	-
144	BRCA2	F	40	12	55	CCI	2	++	-	+
146	BRCA2	F	64	25	80	CMI	3	-	-	+
148	BRCA2	F	62	25	90	CCI et CLI	3	++	-	++
149	BRCA2	F	76	70	60	CCI	2	+++	-	-

**Table 4 : Distribution des caractéristiques des 9 tumeurs du sein mutées pour BRCA2 et des 36 tumeurs du sein BRCAX de l'Institut Bergonié.** *nd non déterminé, F Féminin, M masculin, CCI Carcinome Canalaire Infiltrant, CAI Carcinome Apocrine Infiltrant, CLI Carcinome Lobulaire Infiltrant, CMuc Carcinome Mucineux, CMI Carcinome Médullaire Infiltrant. Score IHC RE et RP (score compris entre 0 et 300 obtenu par la multiplication de l'intensité [1 à 3] avec le pourcentage de cellules positive de [0 à 100]) : score -= ≤20, score += 21-100, score ++= 101-200, score +++= >200. Score IHC ERBB2 selon le score Herceptest™ : score - et + = négatif, score ++= faiblement positif et score +++= fortement positif.*

### 3.2. Identification de profils génomique et transcriptomique spécifiques des tumeurs du sein liées à une mutation constitutionnelle BRCA2

L'analyse transcriptomique dans une série de tumeurs du sein survenant dans un contexte héréditaire comportant 7 tumeurs BRCA2 et 24 tumeurs BRCA1 validée sur 19 tumeurs BRCA2 et 12 tumeurs BRCA1 nous permettent d'identifier des régions génomiques localisées sur les chromosomes 13 et 14 fortement associées aux tumeurs BRCA2 que nous avons étudiées en CGH et locus SNP sur 9 tumeurs BRCA2.

#### 3.2.1. Etablissement d'une signature transcriptomique des tumeurs BRCA2

Les résultats de l'analyse supervisée des données d'expression des tumeurs BRCA2 versus le groupe contrôle BRCA1 identifie 66 gènes différenciant au mieux les tumeurs BRCA2 des tumeurs BRCA1. Les 66 gènes de la signature ont une valeur statistique  $t$ -value ajustée par la correction de Benjamin Hochberg inférieure à 0.01. Ces gènes peuvent être divisés en deux sous-groupes selon la valeur statistique de test  $t$  modérée *Empirical Bayes test*, une valeur  $t$  négative étant associée à la sous-expression du gène dans les tumeurs BRCA2 (22 gènes), et une valeur  $t$  positive étant associée à la surexpression du gène (44 gènes). Dans le cluster de sous-expression, 14 gènes sont localisés sur les bras longs des chromosomes 13 et 14 et particulièrement sur les locus 13q12, 13q14, 13q22, 14q23, 14q24 et 14q32 (Table 5).

L'analyse d'enrichissement de la signature par la méthode GSEA (*Gene Set Enrichment Analysis*) a ensuite confirmé que les gènes localisés sur les chromosomes 13 et 14 sont fortement représentés dans la signature transcriptomique BRCA2 [187]. En effet, sur les 7 premières bandes cytogénétiques concernant les gènes sous-exprimés ayant un score d'enrichissement normalisé le plus fort (NES négatif) dans les tumeurs BRCA2, deux bandes appartiennent au bras long du chromosome 13 (13q13 et 13q14), et trois bandes au bras long du chromosome 14 (14q24, 14q31 et 14q32).

Ainsi, nos résultats montrent que les gènes sous-exprimés différenciant au mieux les tumeurs BRCA2 des tumeurs BRCA1 sont majoritairement situés sur les bras longs des chromosomes 13 et 14 (14 gènes sur 22) sur les bandes cytogénétiques 13q13-14 et 14q24-32.

Bande	Symbole du Gène	Titre du Gène	t	p
13q14	TPT1	tumor protein, translationally-controlled 1	-5.71	0.0047
13q14	WBP4	WW domain binding protein 4	-5.51	0.0048
13q22	PIBF1	progesterone immunomodulatory binding factor 1	-5.36	0.0063
13q12	STARD13	StAR-related lipid transfer domain containing 13	-5.18	0.0071
13q22	COMMD6	COMM domain containing 6	-5.12	0.0074
14q24	C14orf133	VPS33B interacting protein	-6.01	0.0034
14q32	MARK3	MAP/microtubule affinity-regulating kinase 3	-5.74	0.0047
14q23	PPP2R5E	protein phosphatase 2, regulatory subunit B <sup>1</sup> , epsilon	-5.70	0.0047
14q24	ALKBH1	alkB, alkylation repair homolog 1 (E. coli)	-5.69	0.0047
14q32	BTBD6	BTB (POZ) domain containing 6	-5.31	0.0066
14q24	SNW1	SNW domain containing 1	-5.23	0.0070
14q24	ZNF410	zinc finger protein 410	-5.16	0.0071
14q32	PACS2	phosphofurin acidic cluster sorting protein 2	-5.00	0.0082
14q24	TTL5	tubulin tyrosine ligase-like family, member 5	-4.98	0.0082

**Table 5: Liste des gènes du cluster de sous-expression de 22 gènes de la signature BRCA2 localisés sur les chromosomes 13 et 14. Les gènes sont triés selon la valeur du t et leur localisation sur le chromosome 13 ou 14.**

### 3.2.2. Etude des profils génomiques des tumeurs BRCA2

Suite à la caractérisation de gènes sous-exprimés localisés sur les bandes cytogénétiques 13q13-14 et 14q24-32 différenciant au mieux les tumeurs BRCA2 des tumeurs BRCA1, nous avons étudié les profils génomiques de 9 tumeurs BRCA2 au niveau des chromosomes 13 et 14 à l'aide de puces à ADN (4 tumeurs BRCA2 hybridées sur puce *Bacterial Artificial Chromosome* (BAC)-array et 5 autres tumeurs BRCA2 hybridées sur puce SNP-array).

#### 3.2.2.1. Interprétation des profils de puces SNP

L'interprétation des profils de puces SNP repose sur l'analyse conjointe de deux types de données : le statut allélique (fréquence de l'allèle B, *B allele Frequency* ou BAF) et le nombre de copies (intensité du ratio de fluorescence en log<sub>2</sub>, *Log R ratio* ou LRR) (Figure 6).

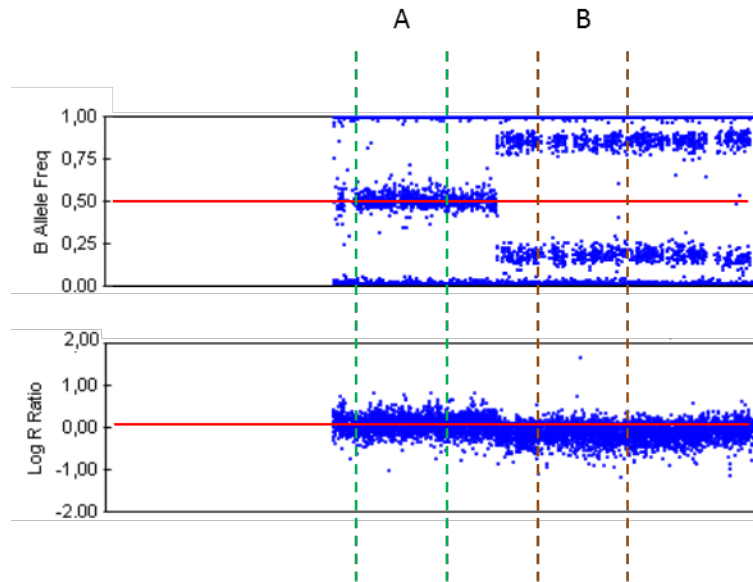
Arbitrairement, les 2 allèles possibles d'un SNP sont désignés « A » et « B ». L'intensité de fluorescence de chaque allèle à un locus est normalisée selon l'intensité de fluorescence d'ADN normaux diploïdes et hétérozygotes.

Le BAF représente le génotype à un locus, sa valeur est corrigée par  $\theta$  qui est donné par la formule suivante:  $\theta = \frac{2}{\pi} \arctan\left(\frac{B}{A}\right)$ , où A et B sont les intensités normalisées des allèles A et B à un locus. La représentation graphique du BAF correspond au pourcentage d'allèle B en fonction de la position génomique des marqueurs. Ainsi, les valeurs du profil BAF sont comprises entre 0 (0% d'allèles B, soit 100% d'allèles A) et 1 (100% d'allèles B, soit 0% d'allèles A). Par conséquent, une valeur de BAF égale à 0.5 signifie une hétérozygotie A/B parfaite se traduisant par 50% d'allèles B et 50% d'allèles A sans déséquilibre allélique. La déviance de la valeur du BAF vers 0 et 1 rend compte d'un déséquilibre allélique et d'une perte d'hétérozygotie appelée également LOH. En pratique, nous observons rarement des valeurs seuils de BAF allant jusqu'aux extrêmes 0 et 1. Ceci peut s'expliquer par la contamination des échantillons tumoraux par des cellules normales, ce qui augmente le nombre de signaux AB hétérozygotes et diminue les valeurs du BAF dans les régions de déséquilibre allélique. De plus, la polyploidie de la cellule entraîne des valeurs de BAF pour des régions hétérozygotes AAB ou ABB par exemple différentes de 0.5.

Le LRR correspond à l'intensité normalisée d'hybridation de l'ADN sur l'axe en  $\log_2$  du ratio de la valeur de R normalisée observée, divisée par la valeur R normalisée attendue, où  $R=A+B$ . La représentation graphique du LRR correspond au  $\log_2$  du ratio du nombre de copies de l'échantillon par rapport à un ADN normal en fonction de la position génomique. Ainsi, 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 du témoin normal est identique. Une augmentation du nombre de copies à un locus se traduit par un  $LRR>0$ , et une diminution par un  $LRR<0$ .

Ainsi, en combinant les analyses, le profil BAF permet d'identifier les LOH, puis la valeur du LRR permet de déterminer si la perte d'hétérozygotie est liée à un gain ( $LRR>0$ ) ou à une perte ( $LRR<0$ ). Cette nouvelle technologie permet à la fois d'être plus résolutif mais aussi d'identifier des réarrangements somatiques complexes comme les isodisomies.





**Figure 6 : Représentation graphique de l'analyse d'un ADN tumoral sur puce SNP, exemple de la tumeur 148. BAF : B Allele Frequency, LRR LogRRatio. Arbitrairement, les 2 allèles possibles d'un SNP sont désignés « A » et « B ». L'intensité de fluorescence de chaque allèle à un locus est normalisée selon l'intensité de fluorescence d'ADN normaux. La région A montre une valeur de BAF égale à 0.5 et une valeur de LRR sensiblement égale à 0, cette région ne présente pas de déséquilibre allélique ni de variation du nombre de copies par-rapport aux ADN contrôles. La région B montre une valeur de BAF égale à 0.8 et une valeur de LRR sensiblement égale à -0.5, cette région présente un déséquilibre allélique et une perte du nombre de copies par-rapport aux ADN contrôles.**

### 3.2.2.2. Identification d'une zone commune de délétion des chromosomes 13 et 14 dans les tumeurs BRCA2

Une approche supervisée utilisant les résultats de profil génomique retrouve de manière spécifique aux tumeurs BRCA2 une association de perte de matériel génétique au niveau des bras longs des chromosomes 13 et 14. Le chevauchement des pertes de matériel nous a conduit à identifier pour les chromosomes 13 et 14, une région commune de délétion présente dans les 9 tumeurs BRCA2 (Figure 7).

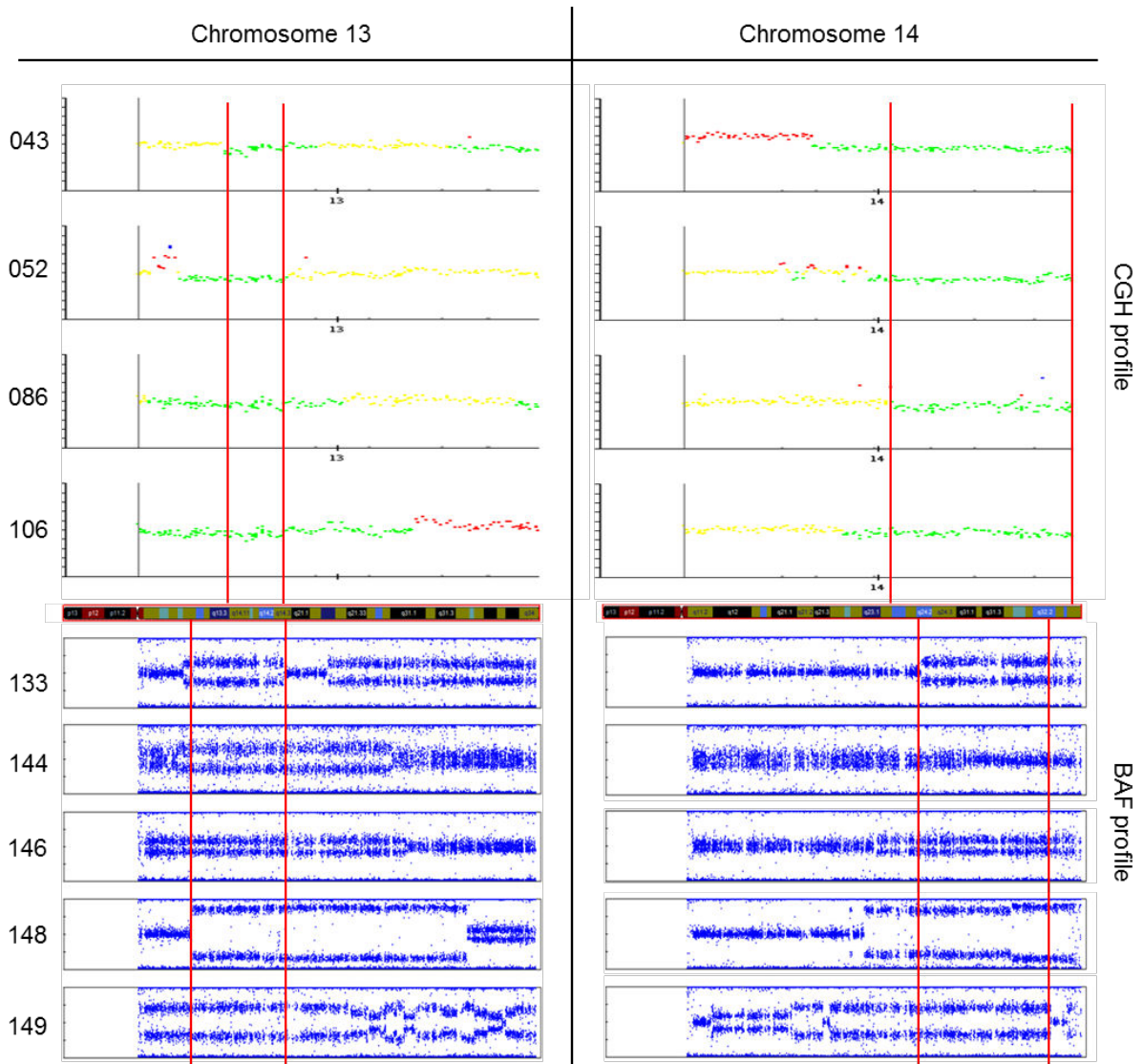
Pour cela, nous avons reporté les coordonnées génomiques en 5' et 3' des régions délétées des tumeurs BRCA2. Pour les tumeurs étudiées en BAC-array (tumeur 43, 52, 86 et 106), les coordonnées génomiques correspondent à la position génomique des clones BACs perdus aux extrémités 5' et 3' des régions de délétion et pour les tumeurs étudiées en puce SNP (tumeurs 133, 144, 146, 148 et 149), les coordonnées génomiques correspondent à la position génomique des marqueurs. Cependant, la résolution de ces deux techniques est différente, car l'espacement entre les 5878 clones de la BAC-array CIT V8 d'Integragen est de 500kb, alors que l'espacement moyen des 620 901 marqueurs de la

Le *Human610-Quad v1.0* d'Illumina est de 4,7kb. La haute résolution des puces SNP a permis de confirmer et d'affiner les observations faites sur les profils BAC-CGH.

La région commune de délétion du chromosome 13 se situe au niveau de la partie proximale du bras long du chromosome mais exclut le gène *BRCA2* pour au moins une tumeur et s'étend entre les bandes chromosomiques 13q13.3 et 13q14.3 (Table 6). Elle comporte 111 gènes dont *TPT1* et *WBP4* qui sont retrouvés dans les 14 gènes sous-exprimés de la signature transcriptionnelle.

En ce qui concerne le chromosome 14, la région commune de délétion est comprise entre les bandes chromosomiques 14q24.2 et 14q32.2 (Table 6) et contient 208 gènes dont *VIPAR*, *ALKBH1*, *SNW1*, *TLL5* et *ZNF410* qui appartiennent également aux 14 gènes sous-exprimés de la signature transcriptionnelle.

L'identification de 2 régions communes de délétion 13q13.3-13q14.3 et 14q24.2-14q32.2 (excluant le gène *BRCA2*) a été confirmée par plusieurs études indépendantes (Table 7). Il est également intéressant de remarquer que les bandes cytogénétiques des chromosomes 13 et 14 les plus représentées dans la signature BRCA2 d'après l'analyse GSEA (13q3-14 et 14q24-q31) comprennent les zones communes de délétion.



**Figure 7 : Profil génomique des chromosomes 13 et 14 des tumeurs BRCA2 en BAC-CGH et en puce SNP.** Les profils génomiques en BAC-CGH ont été obtenus par l'application VAMP de la plateforme CAPweb (Institut Curie), les profils de B Allele Frequency (BAF) ont été visualisés par le logiciel Genome Studio v2010.1. Le tracé jaune des profils en CGH-BAC signifie que le profil est normal, le tracé vert représente une perte, le tracé rouge représente un gain et le tracé bleu représente une amplification. Les barres rouges bornent la zone commune de délétion identifiable à partir des profils génomiques issus des deux techniques CGH et SNP.

Chromosome	Borne 5'		Borne 3'	
13	39.928.703	13q13.3	53.099.075	13q14.3
14	70.418.189	14q24.2	99.851.630	14q32.2

**Table 6 : Zone minimale de délétion des chromosomes 13 et 14 définie à partir de 4 profils CGH-BAC et 5 profils de puces SNP.**

Analyse des profils génomiques de diverses études indépendantes évoquant la perte des chromosomes 13 et 14 des tumeurs BRCA2							
	Tirkkonen 1997	Van Beers 2005	Tischkowitz 2007	Stefansson 2009	Jönsson 2010	Joosse 2012	Pécuchet 2013
<b>Nb de cas BRCA2</b>	15	25	13	18	31	47	29
<b>Borne del chr 13</b>	13cen-q21	13q	13q	<b>13q12.3-14.3</b> 30.622.926- 51.903.522	<b>13q14.2</b> 46.686.707- 47.373.965	<b>13q12-q14</b> 19.072.448- 47.829.251	<b>13q13.1- q21.1</b> 31.331.799- 57.540.925
<b>Borne del chr 14</b>	14q	14q2-3	14q	<b>14q24.2-q24.3</b> 69.875.884- 77.751.255	<b>14q24.3</b> 76.772.977- 77.809.367	<b>14q23.2-32.2</b> 61.994.004- 97.923.256	<b>14q24.2- 32.33</b> 70.422.979- 106.358.708
				<b>14q31.1-31.1</b> 80.450.442- 81.186.435			
				<b>14q31.3-q32.2</b> 87.719.946- 97.210.108			
				<b>14q32.2-q32.2</b> 98.565.152- 98.765.804			

**Table 7: Analyse bibliographique des bornes de délétion des chromosomes 13 et 14 des tumeurs BRCA2.** Tirkkonen et al [145], Van Beers et al [188], Tischkowitz et al [189], Stefansson et al [139], Jönsson et al [143], Joosse et al [144], Pécuchet et al [190].

### 3.2.3. Mise en place d'un test FISH utilisable en routine ciblant les altérations génomiques associées aux tumeurs BRCA2

Nous proposons de mettre au point un test en Fluorescence In Situ Hybridization (FISH) applicable en pratique médicale courante et susceptible d'identifier les tumeurs BRCA2 par délétion des régions d'intérêts des chromosomes 13 et 14. Nous avons sélectionné séparément pour les chromosomes 13 et 14, un BAC complémentaire de la région commune de délétion et un BAC contrôle proche du centromère (Figure 8a). Dans une cellule diploïde pour les chromosomes 13 et 14, la délétion de la région commune de délétion du chromosome 13 et respectivement du chromosome 14, sera matérialisée en FISH par un profil associant un signal de la zone commune de délétion à 2 signaux de la zone péricentromérique (figure 8b et 8c).

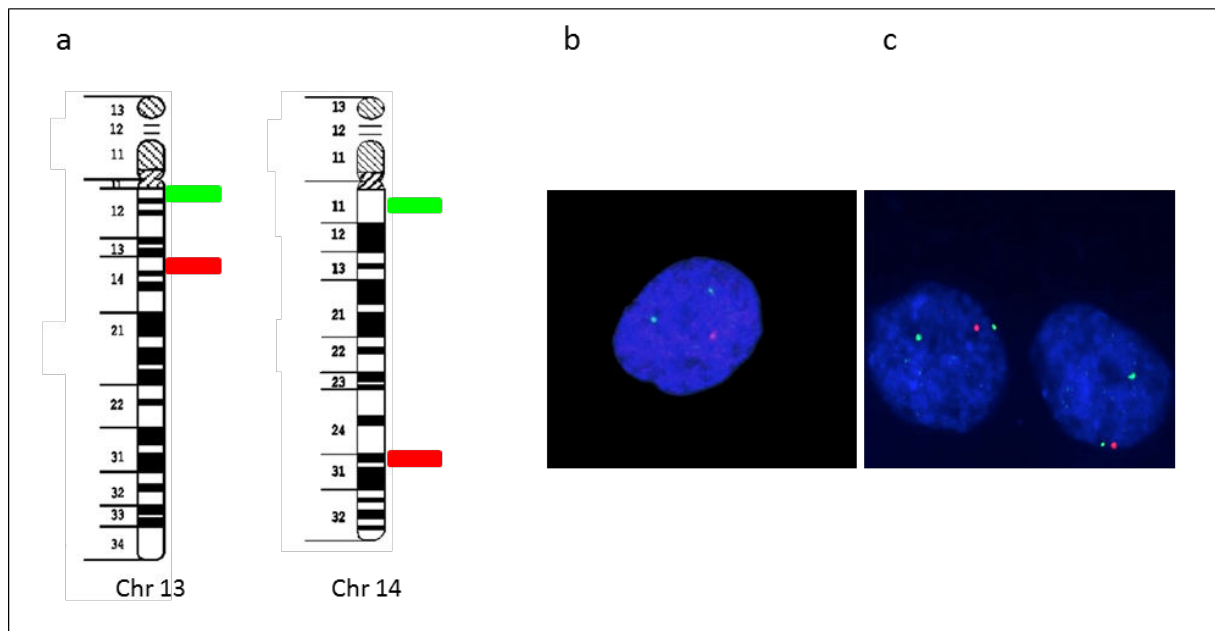
En raison du génome très remanié des tumeurs BRCA2, plusieurs types de profils peuvent être observés. Les noyaux dans lesquels nous observons un déséquilibre entre le nombre de copies de la région d'intérêt et de la région contrôle péricentromérique sont en faveur d'une délétion de la région d'intérêt. En revanche, une ploïdie de 1 signifie qu'aucun déséquilibre entre la région d'intérêt et la région péricentromérique n'est observé, mais que la délétion emporte la région contrôle et la région d'intérêt. Dans ces cellules, les régions étudiées ne sont présentes qu'à une copie.

Les chromosomes 13 et/ou 14 sont considérés « perdus » dans les tumeurs lorsque le pourcentage des noyaux selon les critères énoncés précédemment est supérieur ou égal à 50%. La réalisation de ce test FISH sur 9 tumeurs BRCA2 et 9 Tumeurs BRCA1 sur empreintes congelées après décompte de 50 noyaux évalue la sensibilité et la spécificité de la détection des mutants BRCA2 à 78% et 89% respectivement (Table 8). Il semble que ce test peut être envisagé comme *screening* rapide des délétions des chromosomes 13q et 14q pouvant être associées à une mutation constitutionnelle du gène *BRCA2*.

Le travail présenté a fait l'objet d'une publication dans le journal Plosone « Deletion of chromosomes 13q and 14q is a common feature of tumors with BRCA2 mutations ».

Tumeur	Statut BRCA	Chromosome	Perte signal rouge (2V1R, 2V0R, 3V2R, 4V3R, 5V3R, 4V2R, 4V1R, 6V2R, 3V1R, 1V0R)	Ploidie=1 (1R1V)	Profil équilibré (2R2V, 3R3V, 4R4V, 5R5V)	Gain du signal rouge (2V4R, 0V1R, 1V2R, 1V3R, 2V3R)
A	BRCA2	13q	37	5	8	0
		14q	38	5	7	0
B	BRCA2	13q	3	36	8	3
		14q	0	3	10	37
C	BRCA2	13q	48	0	2	0
		14q	3	2	45	0
D	BRCA2	13q	9	33	6	2
		14q	23	11	16	0
E	BRCA2	13q	0	29	21	0
		14q	38	1	10	1
106	BRCA2	13q	45	0	5	0
		14q	44	0	6	0
133	BRCA2	13q	46	1	3	0
		14q	42	3	5	0
86	BRCA2	13q	41	3	3	3
		14q	44	0	6	0
52	BRCA2	13q	40	2	8	0
		14q	42	3	5	0
F	BRCA2	13q	0	0	50	0
		14q	1	3	46	0
G	BRCA2	13q	1	2	44	3
		14q	0	0	50	0
H	BRCA2	13q	0	45	3	2
		14q	1	42	6	1
I	BRCA2	13q	4	0	44	2
		14q	2	0	40	8
J	BRCA2	13q	0	19	29	2
		14q	0	6	44	0
K	BRCA2	13q	1	0	49	0
		14q	0	6	41	3
L	BRCA2	13q	3	2	45	0
		14q	1	6	35	8
M	BRCA2	13q	3	0	47	0
		14q	0	4	42	4
16	BRCA2	13q	2	5	40	3
		14q	0	0	38	12

**Table 8 : Résultat de l'analyse des zones communes de délétion des chromosomes 13 et 14 sur 9 tumeurs BRCA2 et 9 tumeurs BRCA2. Décompte réalisé sur 50 noyaux. Signal rouge: zone commune de délétion, signal vert : sonde péricentromérique. Les profils observés sont notés xVyR où x et y correspondent au nombre de signaux verts et rouges observés dans un noyau. La colonne « Perte du signal rouge » signifie qu'il y a moins de signaux rouges que de signaux verts et que la région commune de délétion est déléetée par-rapport au signal péricentromérique, la colonne « Ploidie=1 » signifie qu'aucun déséquilibre entre la région d'intérêt et la région péricentromérique n'est observé, mais que la délétion emporte la région contrôle et la région d'intérêt, la colonne « Profil équilibré » signifie qu'il y a autant de signal vert que de signal rouge, il n'y a pas de variation du nombre de copies de la région commune de délétion par-rapport à la région péricentromérique, la colonne « Gain du signal rouge » signifie qu'il y a plus de signaux rouges que de signaux verts et que la région commune de délétion est gagnée par-rapport à la région péricentromérique.**



**Figure 8 : Test FISH susceptible d'identifier les tumeurs BRCA2 par délétion des régions d'intérêt des chromosomes 13 et 14.** (a) Localisation des sondes BAC, signal rouge : zone commune de délétion, signal vert : sonde péricentromérique. FISH réalisée sur une tumeur BRCA2 (Tumeur 52), (b) chromosome 13, (c) chromosome 14. On observe la présence d'un signal rouge pour deux signaux verts, signe de délétion pour un allèle de la zone commune de délétion.

# Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations

Audrey Rouault<sup>1</sup>, Guillaume Banneau<sup>1</sup>, Gaëtan MacGrogan<sup>1,2</sup>, Natalie Jones<sup>1</sup>, Nabila Elarouci<sup>3</sup>, Emmanuelle Barouk-Simonet<sup>4</sup>, Laurence Venat<sup>5</sup>, Isabelle Coupier<sup>6</sup>, Eric Letouze<sup>3</sup>, Aurélien de Reyniès<sup>3</sup>, Françoise Bonnet<sup>1,4</sup>, Richard Iggo<sup>1\*</sup>, Nicolas Sévenet<sup>1,4</sup>, Michel Longy<sup>1,4</sup>

1 French National Institute of Health and Medical Research (INSERM) Unit 916, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France, 2 Pathology Department, Bergonié Cancer Institute, Bordeaux, France, 3 Cartes d'Identité des Tumeurs (CTT) Program, Ligue Nationale Contre Le Cancer, Paris, France, 4 Cancer Genetics Unit, Bergonié Cancer Institute, Bordeaux, France, 5 Department of Medical Oncology, Dupuytren University Hospital, Limoges, France, 6 Cancer Genetics Unit, Val d'Aurelle Regional Cancer Centre, Montpellier, France

## Abstract

**Introduction:** Germline BRCA1 or BRCA2 mutations account for 20–30% of familial clustering of breast cancer. The main indication for BRCA2 screening is currently the family history but the yield of mutations identified in patients selected this way is low.

**Methods:** To develop more efficient approaches to screening we have compared the gene expression and genomic profiles of BRCA2-mutant breast tumors with those of breast tumors lacking BRCA1 or BRCA2 mutations.

**Results:** We identified a group of 66 genes showing differential expression in our training set of 7 BRCA2-mutant tumors and in an independent validation set of 19 BRCA2-mutant tumors. The differentially expressed genes include a prominent cluster of genes from chromosomes 13 and 14 whose expression is reduced. Gene set enrichment analysis confirmed that genes in specific bands on 13q and 14q showed significantly reduced expression, suggesting that the affected bands may be preferentially deleted in BRCA2-mutant tumors. Genomic profiling showed that the BRCA2-mutant tumors indeed harbor deletions on chromosomes 13q and 14q. To exploit this information we have created a simple fluorescence in situ hybridization (FISH) test and shown that it detects tumors with deletions on chromosomes 13q and 14q.

**Conclusion:** Together with previous reports, this establishes that deletions on chromosomes 13q and 14q are a hallmark of BRCA2-mutant tumors. We propose that FISH to detect these deletions would be an efficient and cost-effective first screening step to identify potential BRCA2-mutation carriers among breast cancer patients without a family history of breast cancer.

Citation: Rouault A, Banneau G, MacGrogan G, Jones N, Elarouci N, et al. (2012) Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations. PLoS ONE 7(12): e52079. doi:10.1371/journal.pone.0052079

Editor: Sandra Orsulic, Cedars-Sinai Medical Center, United States of America

Received August 22, 2012; Accepted November 8, 2012; Published December 21, 2012

Copyright: © 2012 Rouault et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was funded by grants from the Charente-Maritime and Pyrenees-Atlantiques committees of the French Cancer League and the Bergerac Lions Club to M.L.; the French National Research Agency (ANR) to R.I.; and a French National Cancer League program grant to M.L. and R.I. This work is part of the national program Cartes d'Identité des Tumeurs (CTT) [http://cit.ligue-cancer.net/] funded and developed by the Ligue nationale contre le cancer. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

\* E-mail: r.iggo@bordeaux.unicancer.fr

## Introduction

Germline mutations in pathways critical for maintenance of genomic integrity confer an increased risk of developing breast cancer [1]. Inherited mutations in two genes, breast cancer 1 (BRCA1) and BRCA2, are associated with a particularly striking increase in breast cancer risk [2]. Consistent with the Knudson two-hit model, both alleles of BRCA1 and BRCA2 are inactivated in tumors, indicating that the genes behave like classic tumor suppressor genes [3]. Their gene products are implicated in the repair of DNA double-strand breaks [4]: BRCA1 is required for recruitment of repair proteins to sites of breakage [5], whereas BRCA2 nucleates RAD51 filament assembly on single-stranded DNA exposed by resection from the break [6]. Loss of these functions leads to genomic instability [7].

The criteria used to select patients for BRCA2 screening are essentially based on the family history. Unfortunately, this approach is wasteful of resources because relatively few familial clusters are caused by germline BRCA2 mutations [8]. This approach also overlooks patients with no overt family history of breast or ovarian cancer who may nevertheless have BRCA2 mutations. Despite numerous efforts, no specific clinical or pathological features have been identified that permit easy identification of BRCA2-associated tumors.

The role BRCA2 plays in repair of double strand breaks by homologous recombination might be expected to give a characteristic pattern of genomic instability but no genomic features have yet been described that can be used to identify these tumors. Gene expression profiling typically places the tumors in the luminal B, high proliferation, estrogen receptor (ER) positive group of the



Table 1. Characteristics of patients and tumors.

ID	Tumor set	BRCA status	Sex	Age at surgery (year)	Tumor size (mm)	Tumor cells (%)	Histologic grade	ER	PR	ERBB2
52	Training	BRCA2	F	35	17	92	3	++	-	-
86	Training	BRCA2	F	46	16	90	3	+++	+	-
106	Training	BRCA2	F	57	22	85	3	+++	+	-
133	Training	BRCA2	F	40	15	75	2	+	+	-
144	Training	BRCA2	F	40	12	55	2	++	-	+
146	Training	BRCA2	F	64	25	80	3	-	-	+
148	Training	BRCA2	F	62	25	90	3	++	-	++
8	Training	BRCAX	F	51	18	90	3	-	-	-
9	Training	BRCAX	F	51	25	95	3	++	++	-
11	Training	BRCAX	F	56	40	78	2	++	+++	-
14	Training	BRCAX	F	45	12	90	2	nd	+++	-
16	Training	BRCAX	F	50	27	95	3	+++	+++	-
22	Training	BRCAX	F	64	18	90	2	+++	+	-
24	Training	BRCAX	F	35	12	70	1	++	-	-
25	Training	BRCAX	F	37	12	92	2	++	+	-
33	Training	BRCAX	F	42	35	73	1	++	-	-
37	Training	BRCAX	F	45	20	92	2	+++	+++	-
38	Training	BRCAX	F	64	13	90	3	+++	-	-
40	Training	BRCAX	F	41	12	95	2	+	++	-
41	Training	BRCAX	F	38	21	92	3	++	+	-
46	Training	BRCAX	F	60	38	90	2	-	-	-
66	Training	BRCAX	F	73	12	90	2	+++	+++	-
75	Training	BRCAX	F	58	14	80	2	-	-	+++
79	Training	BRCAX	F	42	11	90	3	++	++	-
81	Training	BRCAX	F	46	28	80	2	+	++	-
82	Training	BRCAX	F	50	9	85	1	++	++	-
84	Training	BRCAX	F	47	27	92	3	++	+	+
85	Training	BRCAX	F	64	15	90	1	-	+++	-
93	Training	BRCAX	F	44	18	85	2	+++	+++	-
107	Training	BRCAX	F	69	40	80	3	+++	+	-
111	Training	BRCAX	F	73	15	80	1	+++	+	-
3	Validation	BRCAX	F	36	18	95	3	-	+	+++
15	Validation	BRCAX	F	42	15	95	3	-	-	-
17	Validation	BRCAX	F	76	3	95	1	+++	+	-
30	Validation	BRCAX	F	51	nd	95	3	+++	-	nd
48	Validation	BRCAX	F	54	20	90	1	++	++	-
49	Validation	BRCAX	F	49	35	66	2	+++	++	-
65	Validation	BRCAX	F	46	37	95	3	-	-	-
71	Validation	BRCAX	F	43	21	73	2	++	+++	+++
83	Validation	BRCAX	F	50	18	50	2	++	-	-
89	Validation	BRCAX	F	30	30	82	nd	++	+++	-
96	Validation	BRCAX	F	41	25	85	3	-	-	+++
99	Validation	BRCAX	M	63	21	90	1	+++	++	-
43	Genomic	BRCA2	F	38	12	90	2	++	-	-
149	Genomic	BRCA2	F	76	70	60	2	+++	-	-

Footnote. Tumor set: Training set, tumors used to create the gene expression signature; Validation set, BRCAX tumors from Bergonie Cancer Institute; Genomic set, tumors only used for CGH and SNP analysis. nd, not determined. There was no statistically significant difference ( $p > 0.05$ , Fisher test) between the BRCA2 and BRCAX groups for the following comparisons: age at surgery  $<vs \geq 49$  years (median age); tumor size  $<vs \geq 18$  mm (median tumor size); tumor cell content  $<vs \geq 90\%$  (median tumor cell content); ++ vs other ER status; 2 vs other PR status; 2 vs other ERBB2 status.

doi:10.1371/journal.pone.0052079.t001

Stanford classification but this is not specific enough to be useful clinically to identify tumors with BRCA2 mutations [9].

In this study, we have used gene expression and genomic data to identify specific molecular features that distinguish tumors with BRCA2 mutations from tumors with other breast cancer predisposition mutations. Based on these results we have developed a fluorescent in-situ hybridization (FISH) test that can be used to screen for tumors with an increased risk of containing BRCA2 mutations.

## Methods

### Patients and Samples

All samples were from the Bergonie Cancer Institute, Bordeaux, except for sample 144 from the Val d'Aurelle Regional Cancer Center, Montpellier; samples 146 and 148 from the Dupuytren Hospital, Limoges; and the BRCA2 tumors in the validation set from the Curie Institute. The microarray data for the validation set were generously provided by the Translational Research Unit at the Curie Institute, Paris. The control group contained BRCAX tumors, defined as tumors lacking known BRCA1/2 mutations from families with either i) at least three breast cancer-affected first or second-degree relatives; or ii) breast cancer before age 42 or ovarian cancer in two first-degree relatives or two second-degree relatives via a male. 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 authorisation number AC-2008-812, which was approved by the Comité de Protection des Personnes. The BRCA1 and BRCA2 mutation search was made after patients gave signed informed consent in the context of a medical genetic diagnosis of suspected breast cancer predisposition, in compliance with the French law on genetic testing (law number 94-654).

### Tumor and Mutation Characterization

Clinical, pathological and genetic data for each case are listed in Table 1. Immunohistochemistry for ER, progesterone receptor (PR) and HER2 (ERBB2) were performed as previously described [10]. HER2 expression was scored according to the Herceptest system. ER and PR were scored by multiplying the percentage of positive cells by the intensity (score 0-20: -; score 21-100: +; score 101-200: ++; score 201-300: +++). Screening for germline mutations was performed on leucocyte DNA as previously described [10].

### Gene Expression and Genomic Chip Hybridization

RNA was extracted from the tumors as described [10] and hybridized to Affymetrix U133 Plus 2.0 genechip microarrays by the Genopole Alsace-Lorraine genomics platform, except for the validation set which was hybridized by the Curie Institute genomics platform. DNA was extracted from the tumors and hybridized to Integrachip V7 bacterial artificial chromosome (BAC) arrays as described [10]. SNP array profiling was performed on Illumina Human610-Quad v1.0 BeadChips (Illumina, Inc., San Diego, CA) by Integragen (Evry, France). The gene expression and genomic data are available in Array Express under accession numbers E-TABM-854, E-MEXP-3688, E-MEXP-3690 and in GEO under accession number GSE39710.

### Data Processing and Statistical Analyses

Given the rarity of the tumors, it was not possible to avoid processing the tumors in batches; the hybridization dates for the Affymetrix chips are given in the CEL files. The 12 BRCAX

controls for the validation set were chosen because they showed the smallest batch effect relative to the Curie Institute tumors. The 12 BRCAX tumors in the validation set were separate from the 24 BRCAX tumors in the training set. The gene expression data were normalized with the RMA algorithm in R version 2.13.1 [11-13]. To eliminate redundant genes sharing a gene symbol, the most variable probeset was selected based on the standard deviation across the entire dataset. Differentially expressed genes were identified by moderated t-test in limma [14] (an R script for the expression analysis is available on request). The 66 BRCA2 gene signature genes were combined to make a BRCA2 score by summing the mean-centered expression values weighted by the t values from limma. Gene Set Enrichment Analysis (GSEA) was performed with Broad Institute java software [15,16]; the expression dataset was ranked by t-statistic in limma, then enrichment was scored by GSEA for chromosome bands using the MSigDB positional gene sets [15,16]. Centroid-linkage hierarchical clustering was performed in Cluster 3.0 and visualized in TreeView [17]. Array CGH data was normalized with CAPweb software [18] and genomic alterations were visualized with VAMP software using the same thresholds as previously described [10]. SNP data were normalized with Illumina Genome Studio Software v2010.1 using Genotyping module (v1.6.3) and Illumina Genome Viewer module (v1.6.1) to obtain the B Allele Frequency (BAF) for each SNP.

### Fluorescence In Situ Hybridization

To detect deletions on chromosomes 13 and 14, FISH was performed with four BAC probes supplied by BlueGnome (Cambridge, UK). Two clones labeled with SpectrumGreen were used to detect the pericentromeric regions of chromosomes 13 and 14: RP11-408E5 on 13q12.11 (hg19 chr13:19700993-19850551); and RP11-98N22 on 14q11.2 (hg19 chr14:20500968-20660726). Two clones labeled with SpectrumOrange (giving red spots in the figures) were used to detect the deletions on chromosomes 13 and 14: RP11-71C5 on 13q14.11 (hg19 chr13:44921196-45086777) and RP11-242P2 on 14q31.1 (hg19 chr14:80030106-80193689). Nuclei obtained by touch imprints were fixed in 3:1 methanol:acetic acid, washed and dried. The BAC probes were mixed, 5  $\mu$ l of hybridization mix was added per slide, and a coverslip was glued in place to create a hybridization chamber. The sections were denatured at 75°C for 5 minutes and hybridized at 37°C overnight. Stringent washes were performed at 65°C for 10 minutes, then the sections were dehydrated in ethanol and mounted. Images were acquired with a Zeiss Axio Imager Z2 microscope (Göttingen, Germany). The number of red and green spots per nucleus was scored in morphologically intact and non-overlapping nuclei. Deletions were reported when  $\geq 50\%$  of nuclei with the modal number of green spots contained fewer red spots or when they contained single green and red spots.

## Results

### Identification of Genes Differentially Expressed in BRCA2-mutant Tumors

To gain insight into the biology of BRCA2-mutant breast tumors, we performed a supervised analysis looking for genes differentially expressed in BRCA2-mutant and control tumors. All of the tumors came from patients with a familial clustering of breast cancer potentially caused by germline mutation of a breast cancer predisposition gene. The BRCA2-mutant group included 7 tumors from patients with known germline BRCA2 mutations. The control group ("BRCAX") contained 24 patients without mutations in BRCA1 or BRCA2 identifiable by conventional screening. RNA

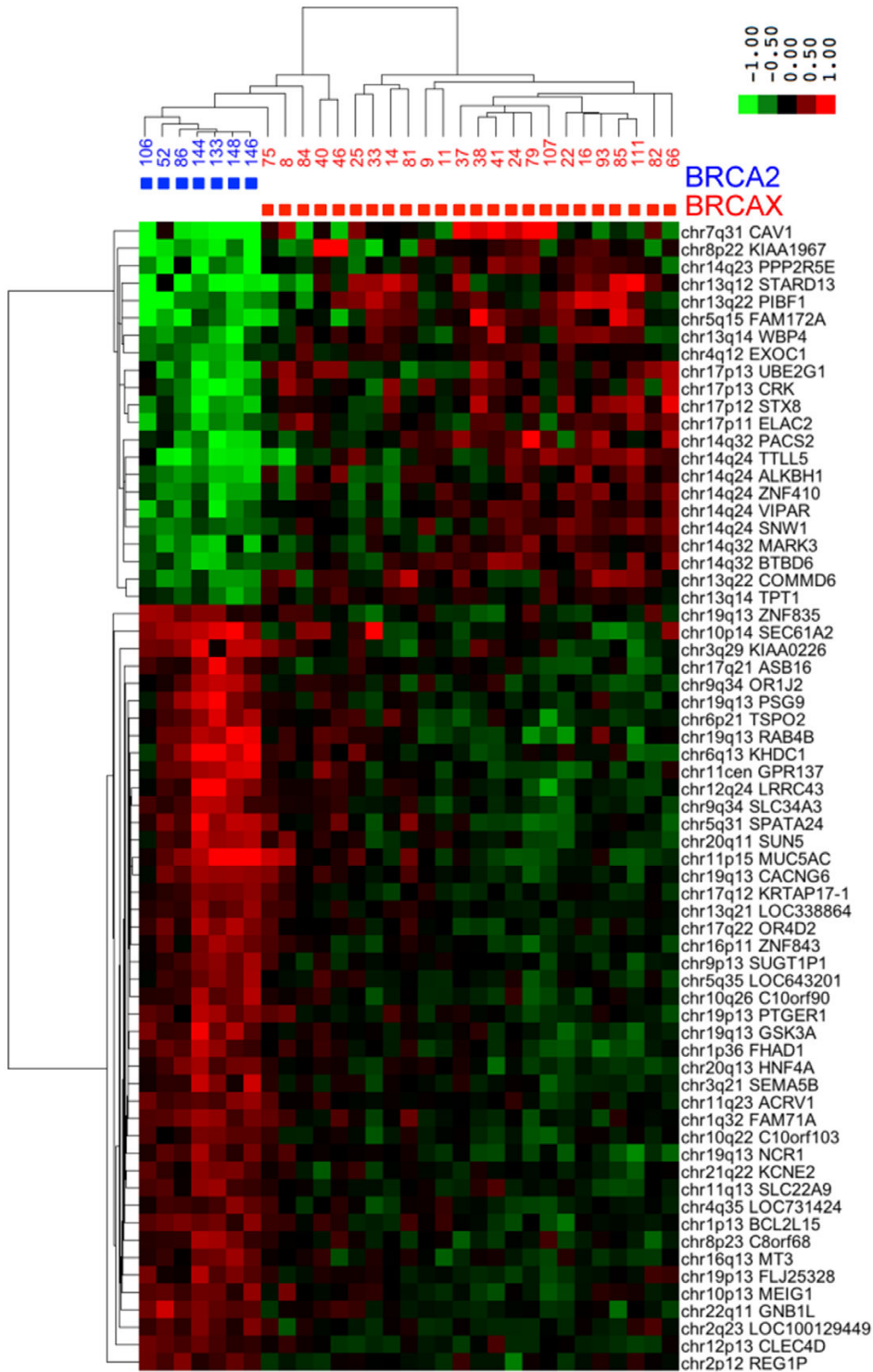


Figure 1. Unsupervised hierarchical clustering of the 66 *BRCA2* signature genes in the training set. There are seven *BRCA2*-mutant tumors and 24 *BRCA2X* tumors (tumors from patients lacking known *BRCA1/2* mutations but with a familial history of breast cancer). The upper left quadrant contains many genes on 13q and 14q that show reduced expression in *BRCA2* tumors.  
doi:10.1371/journal.pone.0052079.g001

Table 2. BRCA2 signature genes.

Affymetrix ID	Gene Symbol	Gene Description	Band	t	p
222127_s_at	EXOC1	exocyst complex component 1	4q12	- 7.05	0.0011
223564_s_at	GNB1L	G protein beta polypeptide 1-like	22q11	6.85	0.0011
632_at	GSK3A	glycogen synthase kinase 3 alpha	19q13	6.42	0.0025
1555377_at	OR4D2	olfactory receptor, family 4, subfamily D, member 2	17q22	6.13	0.0030
208429_x_at	HNF4A	hepatocyte nuclear factor 4, alpha	20q13	6.12	0.0030
207973_x_at	ACRV1	acrosomal vesicle protein 1	11q23	6.21	0.0030
218431_at	C14orf133	VPS33B interacting protein	14q24	- 6.01	0.0034
1552510_at	SLC34A3	solute carrier family 34 (sodium phosphate), member 3	9q34	5.9	0.0041
204690_at	STX8	syntaxin 8	17p12	- 5.84	0.0044
227630_at	PPP2R5E	protein phosphatase 2, regulatory subunit B9 epsilon	14q23	- 5.7	0.0047
205621_at	ALKBH1	alkB, alkylation repair homolog 1 (E. coli)	14q24	- 5.69	0.0047
202569_s_at	MARK3	MAP/microtubule affinity-regulating kinase 3	14q32	- 5.74	0.0047
216520_s_at	TPT1	tumor protein, translationally-controlled 1	13q14	- 5.71	0.0047
230055_at	KHDC1	KH homology domain containing 1	6q13	5.6	0.0048
221966_at	GPR137	G protein-coupled receptor 137	11cen	5.62	0.0048
207733_x_at	PSG9	pregnancy specific beta-1-glycoprotein 9	19q13	5.59	0.0048
1555614_at	SUGT1P1	suppressor of G2 allele of SKP1 (S. cerevisiae) pseudogene 1	9p13	5.57	0.0048
1552772_at	CLEC4D	C-type lectin domain family 4, member D	12p13	5.57	0.0048
203598_s_at	WBP4	WW domain binding protein 4 (formin binding protein 21)	13q14	- 5.51	0.0048
1563639_a_at	FHAD1	forkhead-associated (FHA) phosphopeptide binding domain 1	1p36	5.54	0.0048
234680_at	KRTAP17-1	keratin associated protein 17-1	17q12	5.52	0.0048
1562657_a_at	C10orf90	chromosome 10 open reading frame 90	10q26	5.45	0.0055
236979_at	BCL2L15	BCL2-like 15	1p13	5.39	0.0061
221095_s_at	KCNE2	potassium voltage-gated channel, Isk-related family, member 2	21q22	5.4	0.0061
213239_at	PIBF1	progesterone immunomodulatory binding factor 1	13q22	- 5.36	0.0063
1567257_at	OR1J2	olfactory receptor, family 1, subfamily J, member 2	9q34	5.34	0.0064
225389_at	BTBD6	BTB (POZ) domain containing 6	14q32	- 5.31	0.0066
207778_at	REG1P	regenerating islet-derived 1 pseudogene	2p12	5.3	0.0066
226005_at	UBE2G1	ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, yeast)	17p13	- 5.25	0.0070
215424_s_at	SNW1	SNW domain containing 1	14q24	- 5.23	0.0070
1564112_at	FAM71A	Family with sequence similarity 71, member A	1q32	5.25	0.0070
237980_at	LINC00347	hypothetical LOC338864	13q21	5.24	0.0070
213103_at	STARD13	StAR-related lipid transfer (START) domain containing 13	13q12	- 5.18	0.0071
237257_at	RAB4B	RAB4B, member RAS oncogene family	19q13	5.19	0.0071
201767_s_at	ELAC2	elaC homolog 2 (E. coli)	17p11	- 5.2	0.0071
209944_at	ZNF410	zinc finger protein 410	14q24	- 5.16	0.0071
1558641_at	SPATA24	spermatogenesis associated 24	5q31	5.2	0.0071
212735_at	KJAA0226	Beclin-1 associated RUN domain containing protein	3q29	5.17	0.0071
215449_at	TSPO2	translocator protein 2	6p21	5.15	0.0071
1553253_at	ASB16	ankyrin repeat and SOCS box-containing 16	17q21	5.14	0.0071
231625_at	SLC22A9	solute carrier family 22 member 9	11q13	5.2	0.0071
225312_at	COMMD6	COMM domain containing 6	13q22	- 5.12	0.0074
217187_at	MUC5AC	mucin 5AC, oligomeric mucus	11p15	5.1	0.0077
1553728_at	LRRC43	leucine rich repeat containing 43	12q24	5.07	0.0079
1552863_a_at	CACNG6	calcium channel, voltage-dependent, gamma subunit 6	19q13	5.07	0.0079
217095_x_at	NCR1	natural cytotoxicity triggering receptor 1	19q13	5.06	0.0079
223610_at	SEMA5B	semaphorin 5b	3q21	5.06	0.0079
203065_s_at	CAV1	caveolin 1, caveolae protein, 22 kDa	7q31	- 5.03	0.0080
202226_s_at	CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)	17p13	- 5.04	0.0080

Table 2. Cont.

Affymetrix ID	Gene Symbol	Gene Description	Band	t	p
235416_at	LOC643201	centrosomal protein 192 kDa pseudogene	5q35	5.03	0.0080
1557827_at	C10orf103	chromosome 10 open reading frame 103	10q22	5.03	0.0080
225187_at	KIAA1967	DBC1 deleted in breast cancer 1	8p22	- 4.98	0.0082
212936_at	FAM172A	family with sequence similarity 172, member A	5q15	- 4.99	0.0082
215898_at	TTL5	tubulin tyrosine ligase-like family, member 5	14q24	- 4.98	0.0082
212778_at	PACS2	phosphofurin acidic cluster sorting protein 2	14q32	- 5	0.0082
1562914_a_at	FLJ25328	hypothetical LOC148231	19p13	5	0.0082
215826_x_at	ZNF835	zinc finger protein 835	19q13	4.97	0.0084
238158_at	MEIG1	meiosis expressed gene 1 homolog (mouse)	10p13	4.97	0.0084
219499_at	SEC61A2	Sec61 alpha 2 subunit ( <i>S. cerevisiae</i> )	10p14	4.94	0.0087
207650_x_at	PTGER1	prostaglandin E receptor 1 (subtype EP1), 42 kDa	19p13	4.94	0.0087
237188_x_at	SUN5	Sad1 and UNC84 domain containing 5	20q11	4.92	0.0091
1557679_at	C8orf68	chromosome 8 open reading frame 68	8p23	4.91	0.0092
224256_at	LOC100129449	PRO2055	2q23	4.89	0.0095
1564362_x_at	ZNF843	zinc finger protein 843	16p11	4.88	0.0097
205970_at	MT3	metallothionein 3	16q13	4.87	0.0098
1569095_at	LOC731424	hypothetical LOC731424	4q35	4.87	0.0098

Footnote. t: moderated t-statistic for 66 genes that best discriminate between BRCA2 and BRCAX tumors. p: p-value after Benjamini Hochberg correction (all genes had an unadjusted p-value < 0.0001).

doi:10.1371/journal.pone.0052079.t002

from these 31 tumors was tested on Affymetrix gene expression chips. Sixty-six genes were differentially expressed in the BRCA2 and BRCAX groups at a false discovery rate < 0.01 after Benjamini Hochberg correction for multiple testing (Table 2). Hierarchical clustering confirmed, as expected, that the differentially expressed genes clearly split the tumors into two groups (Figure 1). The BRCA2 group in the heatmap contains five BRCAX tumors that may represent tumors whose BRCA2 mutations were missed by screening or tumors that phenocopy BRCA2.

#### Validation of a Putative BRCA2 Signature

We combined the differentially expressed genes in Table 2 to make a potential BRCA2 gene expression signature. Receiver operating characteristic (ROC) analysis showed that the area under the curve (AUC) for classification of the training set was 1.0 with the BRCA2 signature genes, indicating perfect classification of the tumors. This is not surprising given the small size of the dataset. To test for overfitting, we analyzed an independent validation set of 19 BRCA2-mutant tumors from the Curie Institute genetics clinic and 12 BRCAX from the Bergonie Cancer Institute. Given the rarity of the disease it is unfortunately difficult to avoid batch effects that might confound the result. Nevertheless, the AUC of the ROC curve was 0.76 in the validation set (Figure 2), indicating that the BRCA2 signature was able to classify BRCA2-mutant tumors reasonably well. Hierarchical clustering confirmed that the BRCA2 signature genes were differentially expressed in the validation set (Figure 3). While this suggests that the BRCA2 signature has discriminant value in our tumors and in the validation set from the Curie Institute we note that this is not generally the case because the signature does not identify BRCA2-mutant tumors in some published datasets. For example, the AUC in the Waddell dataset [19] was 0.64, perhaps because of differences in the technology or in the populations studied. We conclude that the BRCA2 signature may have discriminant value in tumors processed according to our protocol.

#### Gene Set Enrichment Analysis (GSEA) Reveals the Mechanism Behind the BRCA2 Signature

The striking feature of the heatmap in Figure 1 is the cluster of 22 genes showing reduced expression in BRCA2-mutant tumors.

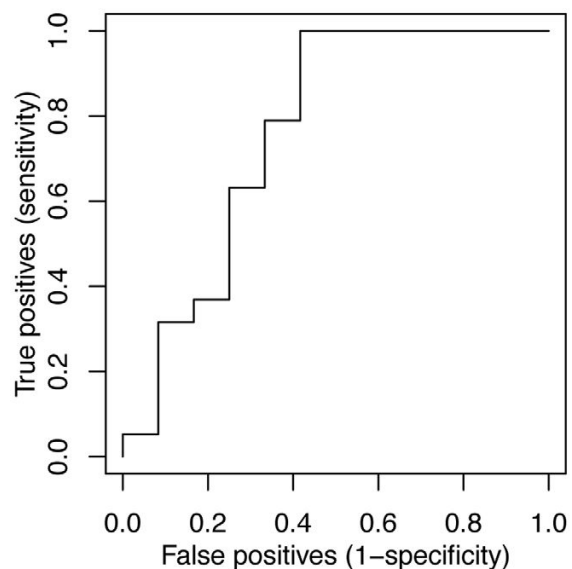


Figure 2. ROC analysis of the BRCA2 signature in the validation set. Each tumor was given a score that was a weighted sum of the mean centered gene expression levels for each gene in the signature. The validation set contained 19 BRCA2 and 12 BRCAX tumors. The AUC was 0.76.

doi:10.1371/journal.pone.0052079.g002

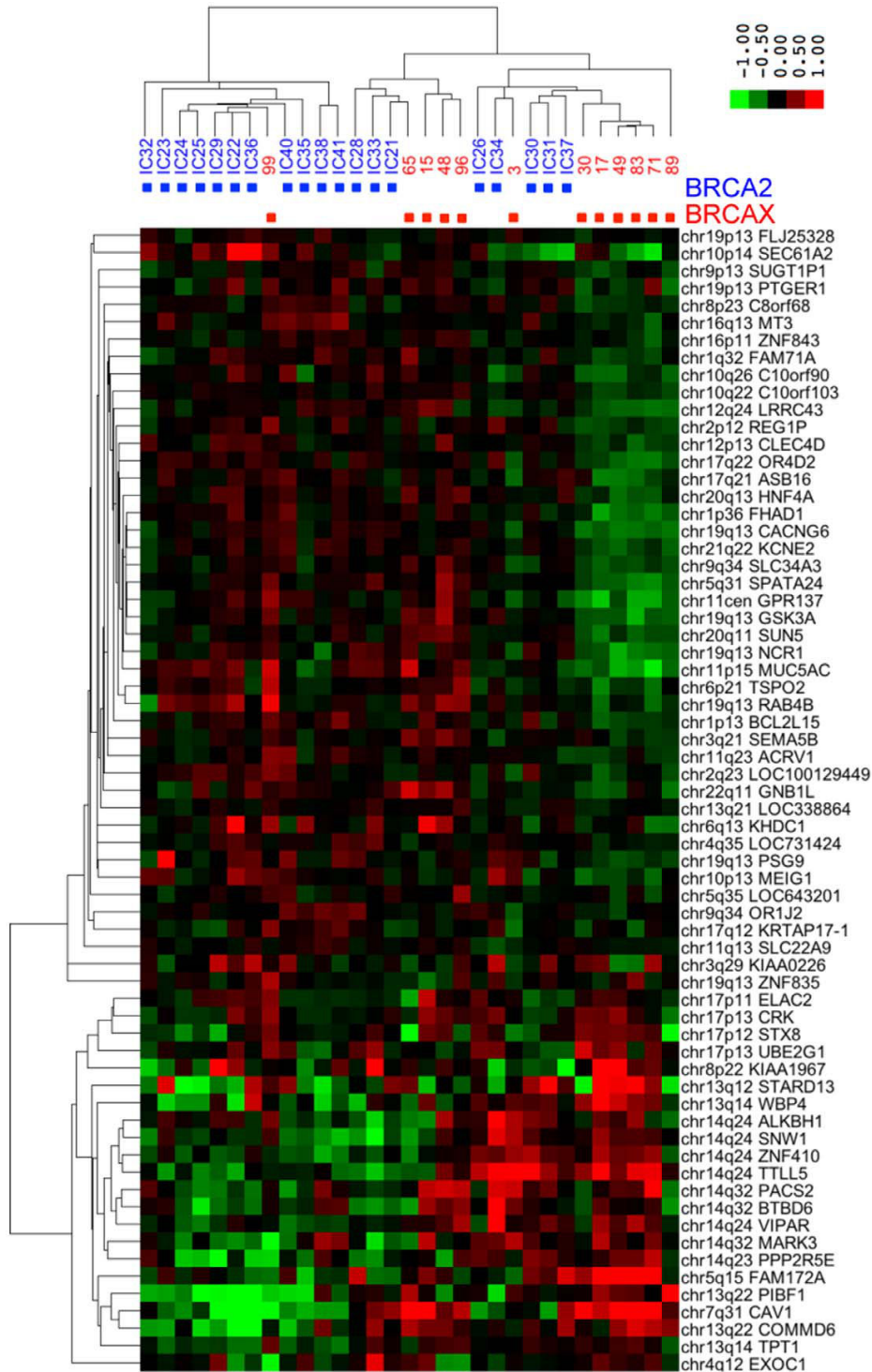


Figure 3. Unsupervised hierarchical clustering of the 66 *BRCA2* signature genes in the validation set. There are 19 *BRCA2*-mutant tumors and 12 *BRCAX* tumors. The lower left quadrant contains many genes on 13q and 14q that show reduced expression in *BRCA2* tumors. doi:10.1371/journal.pone.0052079.g003

Table 3. GSEA for loss of chromosomal bands.

Band	Genes	ES	NES
13q14	67	- 0.63	- 2.75
14q31	22	- 0.81	- 2.71
13q13	22	- 0.74	- 2.45
14q24	77	- 0.54	- 2.43
17p13	185	- 0.44	- 2.3
14q32	105	- 0.48	- 2.28
10q26	72	- 0.51	- 2.27
4p16	91	- 0.49	- 2.25

Footnote. The genes column shows the number of genes used to score the band. The nominal, FDR and FWER p-values were all <0.001. ES, enrichment score; NES normalized enrichment score.  
doi:10.1371/journal.pone.0052079.t003

These genes show a correlation of 0.90 in the heatmap. To exclude fortuitous hybridization as an explanation for this strong clustering we verified that the probe sequences were different and that they were labeled by Affymetrix as valid, non-cross-hybridizing probes for the indicated genes. Fourteen of the 22 BRCA2 signature genes showing reduced expression are from

chromosomes 13 and 14. To determine whether this was due to chance, we ranked the dataset by moderated t statistic (BRCA2 vs control), then performed GSEA with gene sets derived from individual chromosomal bands. The bands most frequently lost are shown in Table 3. The enrichment for bands on 13q and 14q was highly significant ( $p < 0.001$  for the family-wise error rate, the most stringent criterion in the Broad Institute implementation of GSEA). The most likely explanation for underexpressed genes to be derived from specific chromosomal bands is deletion of those bands in the corresponding tumors.

#### CGH and SNP Analysis of BRCA2-mutant Tumors

To test directly for loss of the regions containing the BRCA2 signature genes we measured DNA copy number on CGH and SNP chips. The resulting CGH and SNP profiles confirmed that the incriminated regions are indeed deleted in the BRCA2-mutant tumors (Figure 4). The common region of overlap of the deletions extends from 13q13.3 to 13q14.3 and from 14q24.2 to 14q32.2. The cumulative rates of gain and loss for the BRCA2 and BRCAX tumors are shown in Figure 5. This shows that the long arms of both chromosomes 13 and 14 contain large regions that are preferentially deleted in the BRCA2-mutant tumors. We conclude that the BRCA2 signature genes are differentially expressed because they are deleted in the BRCA2 tumors.

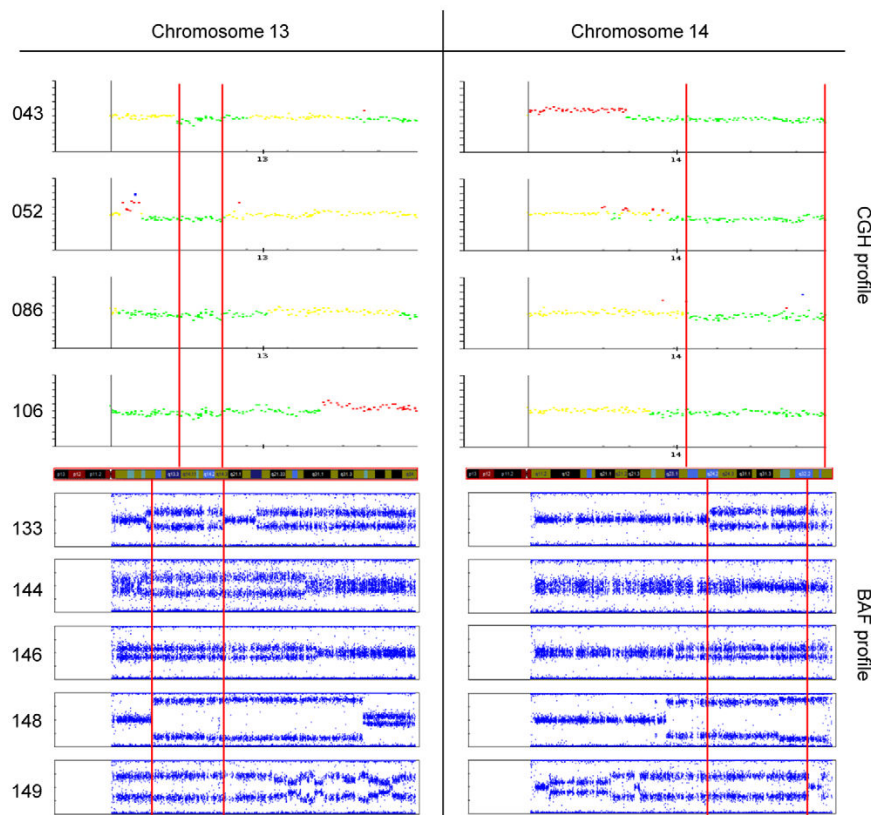


Figure 4. Genomic profiles in the training set. Upper panels: BAC-CGH profiles of BRCA2-mutant tumors showing gains in red, losses in green and modal copy number in yellow. Lower panels: BAF profiles of BRCA2-mutant tumors on Illumina SNP arrays. The boundaries of the common regions of deletion on chromosomes 13 and 14 are marked by vertical red lines.  
doi:10.1371/journal.pone.0052079.g004

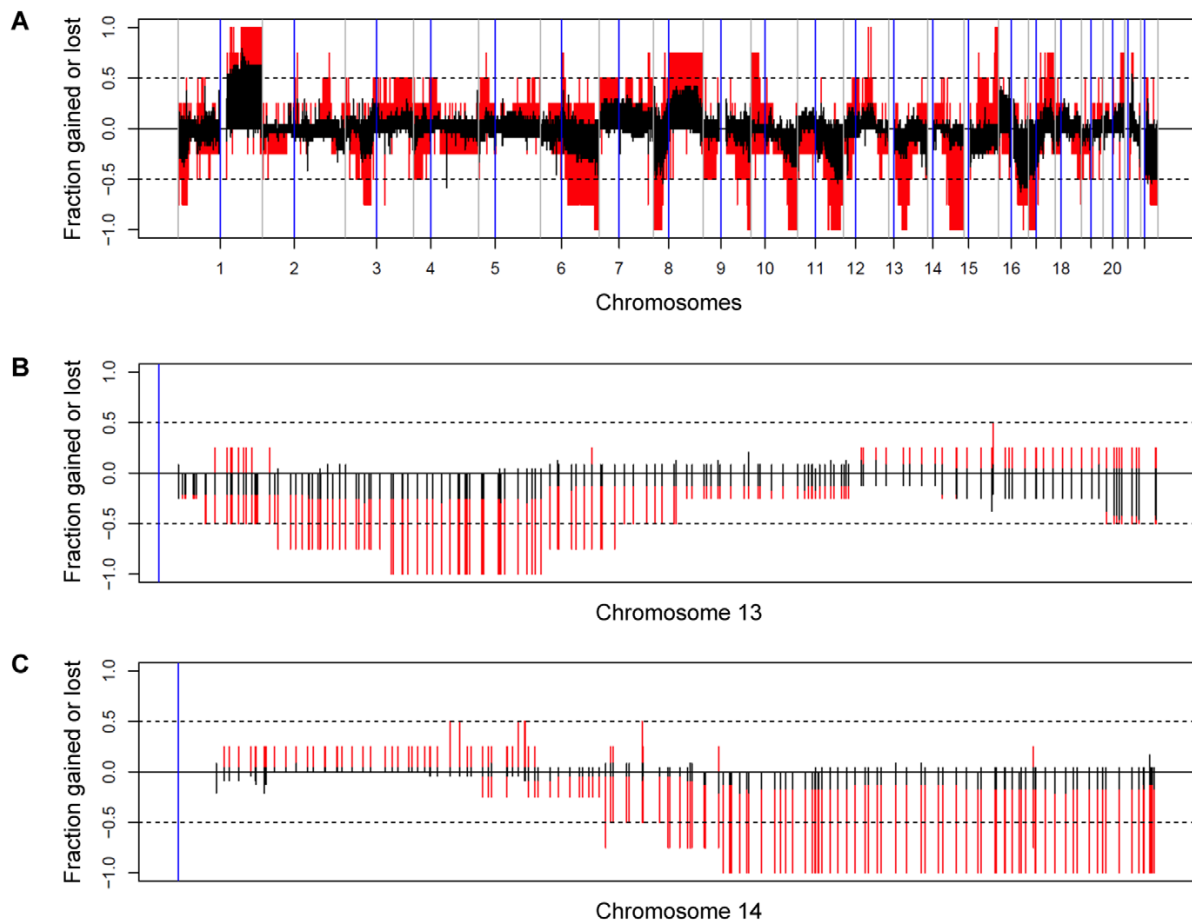


Figure 5. Cumulative rates of gain and loss for tumors analyzed by CGH (red, 4 *BRCA2*-mutant tumors; black, 24 *BRCA2*-wildtype tumors). A, All chromosomes; B, Chromosome 13; C, Chromosome 14. Each vertical line in B & C corresponds to an individual BAC probe. When the red line reaches 2.0, all of the tumors showed loss for that probe. doi:10.1371/journal.pone.0052079.g005

#### Identification of Deletions by FISH

If the signature works by detecting large deletions on chromosomes 13 and 14, it would be better to screen tumors in a clinical setting by FISH rather than by gene expression or CGH/SNP profiling. FISH is ideally suited to detecting small changes in copy number. To test whether it would be feasible to screen for *BRCA2*-mutant tumors in this way, we performed FISH with probes mapping to the regions commonly deleted on chromosomes 13 and 14 (Figure 6). We tested nine *BRCA2* tumors and nine control *BRCA2*-wildtype tumors, of which five *BRCA2* and eight *BRCA2*-wildtype were not previously characterized by CGH. The results are expressed as the percentage of nuclei with less than the modal number of spots for the centromeric probes or with a ploidy of one for both probes (Table 4). The tumors were scored as “loss” when the percentage was  $\geq 50\%$ , and “other” when it was  $< 50\%$ . Contingency tables for the chromosomes individually or for both chromosomes together are shown in Table 5. For both chromosomes scored together, the sensitivity and specificity for detection of *BRCA2*-mutant tumors were 78% and 89%, respectively. We conclude that FISH provides a simple technique to screen tumors for deletions on 13q and 14q that may be associated with *BRCA2* mutations.

#### Discussion

The main conclusion from our study is that deletions on chromosomes 13q and 14q are a common feature of *BRCA2*-mutant tumors. We initially set out to identify a gene expression signature that would distinguish these tumors from other tumors in patients presenting to our genetics clinics. Hierarchical clustering of the genes in the signature split the tumors into two groups in both the training and the validation sets, suggesting that the signature detects a signal that is useful for classification of the tumors. Given the GSEA and SNP/CGH results we strongly suspect that the reduced expression of the genes in the signature is caused by a reduction in the DNA copy number of the deleted regions. It is more difficult to detect deletion than amplification in gene expression data, because the former may further decrease a barely detectable signal whereas the latter can increase expression 100-fold. This probably explains why the genes in the signature are a minority of the genes in the deleted regions. Given the difficulty in measuring weakly expressed genes it is not surprising that previously reported *BRCA2* gene expression signatures did not highlight deletion of chromosomes 13 and 14 as a potential discriminating factor [19,20]. In contrast, deletion of these regions



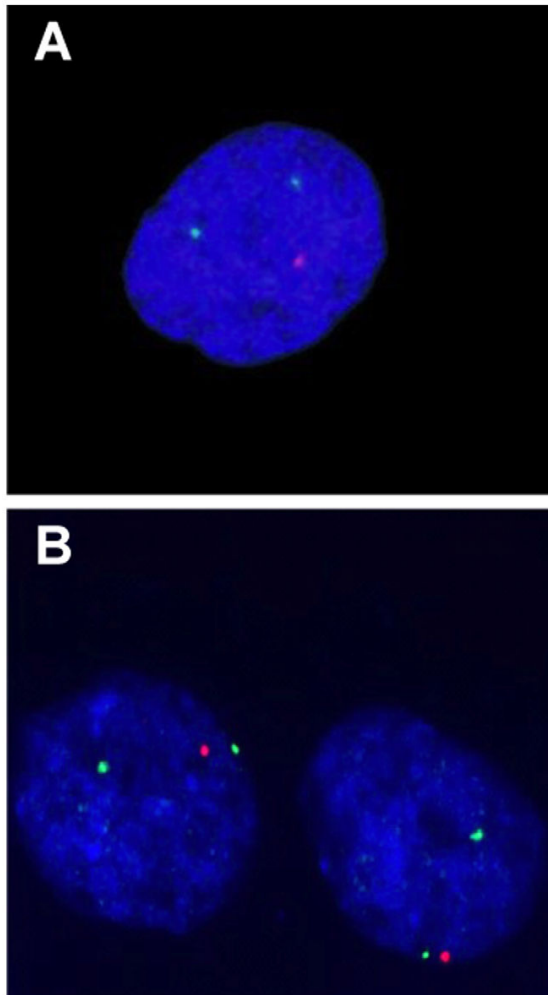


Figure 6. FISH with probes in the region of common deletion in a *BRCA2*-mutant tumor. A, chromosome 13; B, chromosome 14. Red: probe in the deleted region; Green, pericentromeric probe. Each nucleus contains two green spots and one red spot, indicating that the tumor is diploid for chromosomes 13 and 14 but has heterozygous deletions in the regions tested by the red probes. doi:10.1371/journal.pone.0052079.g006

was noted in several previous DNA copy number and SNP studies [7,21–25]. In addition to published studies, we examined the GISTIC database (Tumorscape Release 1.6) [26] to determine whether loss of chromosomes 13 and 14 is a common event in breast cancer. Several regions are reported as harboring deletions on chromosome 13 (hg18 chr13:44680312–57088104, 57088104–114059427, 18097312–46301361 and 50901262–114059427), as expected given the presence of *BRCA2* and *RB1* on 13 q. In contrast, GISTIC reports no regions as being deleted on chromosome 14 in breast cancer at above the background rate ( $q > 0.25$ ).

There are several possible explanations for selective deletion of specific genomic regions in *BRCA2* tumors. The commonly deleted region on chromosome 13 is distal to the *BRCA2* gene, but we can not altogether exclude that *BRCA2* itself may be a driver gene in

Table 4. FISH with probes in the region of common deletion on chromosomes 13 and 14.

ID	BRCA status	chr 13	chr 14
52	BRCA2	84	89
86	BRCA2	90	86
106	BRCA2	100	87
133	BRCA2	93	89
A	BRCA2	84	83
B	BRCA2	100	0
C	BRCA2	87	7
D	BRCA2	100	62
E	BRCA2	100	73
16	BRCA2	0	0
F	BRCA2	0	2
G	BRCA2	0	0
H	BRCA2	100	100
I	BRCA2	4	0
J	BRCA2	0	0
K	BRCA2	2	0
L	BRCA2	7	3
M	BRCA2	10	0

Footnote. The table shows the percentage of nuclei with less than the modal ploidy or with ploidy=1 for both the centromeric and the deletion probes. Tumours A-M were not characterized by CGH. doi:10.1371/journal.pone.0052079.t004

some cases, for example if there were complex genomic rearrangements on 13 q. *BRCA2* was not part of the gene signature, probably because the Affymetrix probes for *BRCA2* are not sensitive enough (the measured level was close to background and showed minimal variation). The best reporters for copy

Table 5. Contingency table summarizing the FISH data for deletions on chromosomes 13 and 14.

Chr 13 and 14	Other	Loss
BRCA2	2	7
BRCAX	8	1
p=0.015		
Chr 13	Other	Loss
BRCA2	0	9
BRCAX	8	1
p=0.0004		
Chr 14	Other	Loss
BRCA2	2	7
BRCAX	8	1
p=0.015		

Footnote. "Loss" refers to cases where  $\geq 50\%$  of nuclei had less than the modal ploidy or had ploidy=1. "Other" refers to cases where the value was  $< 50\%$ . The p value is for a Fisher exact test. The values for "Chr13 and 14" refer to cases where both chromosomes were affected. doi:10.1371/journal.pone.0052079.t005

number are housekeeping genes that lack feedback or exogenous regulation. By their nature these genes shed no light on the mechanism driving deletion. An alternative explanation is that loss of BRCA2 function generates repair intermediates or triggers checkpoint responses that are toxic in the presence of specific genes located in the deleted regions. Loss of these genes would allow the cell to resume division and form a tumor. This model predicts that the driver genes in the deleted regions should be DNA repair or checkpoint genes. ALKBH1 could have this effect, but few other genes in the BRCA2 signature are obvious candidates for these roles. Another possibility is that the deleted regions contain fragile sites that are more difficult to repair in the absence of BRCA2. Fragile sites are prone to replication fork collapse, a process that often leads to the formation of double strand breaks that require repair by homologous recombination. BRCA2 is required for loading of RAD51 to initiate homologous recombination [6] so increased breakage at fragile sites in the affected regions is certainly a possibility.

Screening for BRCA2 mutations is widely performed in genetics laboratories to explain familial clustering of breast cancer. Our study design focused on patients referred to genetics clinics because this is the context in which the need to distinguish BRCA2-mutant from other tumors most commonly arises. Because of the size of the BRCA2 gene it can take many months to identify mutations. This is rarely a problem in the context of genetic counseling because some interventions can be undertaken without knowledge of the mutation (for example, more frequent screening with imaging techniques) and others may even benefit from the delay by giving patients more time for reflection (for example, prophylactic mastectomy and oophorectomy). The same can not be said of medical treatment of established tumors, which must be delivered without delay. The advent of medical treatments specific for BRCA2-mutant tumors has created a need to identify these tumors on a more rapid time scale than has hitherto been considered necessary. In particular, BRCA2 defects are synthetic lethal with inhibition of poly-ADP-ribose polymerase 1 (PARP1) [27,28]. We note that the BRCA2 group in the training set contains five BRCA2 tumors which presumably either phenocopy BRCA2 mutation or contain BRCA2 mutations that evaded detection by sequencing. It would be interesting to know whether tumors that phenocopy BRCA2 mutation are also sensitive to PARP inhibitors.

In the long term it is likely that diagnostic laboratories will routinely use next generation sequencing (NGS) to identify

mutations in BRCA2 and other relevant genes in the diagnostic biopsy when the patient initially presents with cancer. This is technically feasible but rarely performed outside major centers at present because of the cost and the complexity of the downstream bioinformatic analysis. To bridge the gap while waiting for NGS to become more widely available we propose to use FISH to screen breast tumors for deletions on 13q and 14q in order to identify tumors potentially associated with BRCA2. The technology for FISH is very well established for diagnosis of ERBB2 amplification in sporadic breast tumors. It would require only a small modification of existing protocols to screen for loss of 13q and 14q in centers that already screen for ERBB2 amplification by FISH. Patients whose tumors harbor deletions in those regions could then be screened by sequencing to identify either germline or somatic BRCA2 mutations, followed by treatment with PARP inhibitors, if appropriate.

### Conclusion

We have shown that breast tumors arising in patients with germline BRCA2 mutations have a higher frequency of deletions on 13q and 14q than is seen in other breast tumors. We propose that FISH for deletions on these chromosomes would be a rapid and technically feasible first step to enrich for tumors worth screening for BRCA2 mutations. This would greatly facilitate the selection of patients for PARP inhibitor therapy.

### Acknowledgments

We would like to thank the patients who contributed samples to the study. We thank Véronique Fermeaux, Frédéric Bibeau and the Bergonie Cancer Institute Biological Resources Center for providing samples. We thank Marc-Henri Stern for helpful discussions and Thierry Dubois, Marc-Henri Stern and the Translational Research Unit at the Curie Institute for providing the microarray data for the validation set. We thank the members of the Affymetrix expression array platform at IGBMC, Strasbourg.

### Author Contributions

Conceived and designed the experiments: ML RI NS. Performed the experiments: AR GB. Analyzed the data: GM RI AR EL AdR NE. Contributed reagents/materials/analysis tools: EBS ML LV IC. Wrote the paper: RI ML AR. Performed mutation screening: NJ FB NS.

### References

- Walsh T, King MC (2007) Ten genes for inherited breast cancer. *Cancer Cell* 11: 103–105.
- Fackenthal JD, Olopade OI (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* 7: 937–948.
- Collins N, McManus R, Wooster R, Mangion J, Seal S, et al. (1995) Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12–13. *Oncogene* 10: 1673–1675.
- Boulton SJ (2006) Cellular functions of the BRCA tumour-suppressor proteins. *Biochem Soc Trans* 34: 633–645.
- Greenberg RA, Sobhian B, Pathania S, Cantor SB, Nakatani Y, et al. (2006) Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. *Genes Dev* 20: 34–46.
- Yang H, Li Q, Fan J, Holloman WK, Pavletich NP (2005) The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA–ssDNA junction. *Nature* 433: 653–657.
- Stefansson OA, Jonasson JG, Johannsson OT, Olafsdottir K, Steinarsdottir M, et al. (2009) Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Res* 11: R47.
- Moller P, Hagen AI, Apold J, Maehle L, Clark N, et al. (2007) Genetic epidemiology of BRCA mutations—family history detects less than 50% of the mutation carriers. *Eur J Cancer* 43: 1713–1717.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869–10874.
- Banneau G, Guedj M, Macrogan G, de Mascarel I, Velasco V, et al. (2010) Molecular apocrine differentiation is a common feature of breast cancer in patients with germline PTEN mutations. *Breast Cancer Res* 12: R63.
- R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>. Accessed 2012 Nov 22.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barday YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264.
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, et al. (2003) PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267–273.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–14868.

18. CAPweb. Copy number Array analysis Platform on the web: A web-based platform for copy number array management and analysis, version 2.4. Available: <http://bioinfo-out.curie.fr/CAPweb/>. Accessed 2012 Nov 22.
19. Waddell N, Arnold J, Cocciardi S, da Silva L, Marsh A, et al. (2010) Subtypes of familial breast tumours revealed by expression and copy number profiling. *Breast Cancer Res Treat* 123: 661–677.
20. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, et al. (2001) Gene expression profiles in hereditary breast cancer. *N Engl J Med* 344: 539–548.
21. Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, et al. (2005) Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. *Cancer Res* 65: 7612–7621.
22. Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Holm K, et al. (2010) Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* 12: R42.
23. Joosse SA, Brandwijk KI, Devilee P, Wesseling J, Hogervorst FB, et al. (2012) Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. *Breast Cancer Res Treat* 132: 379–389.
24. Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, et al. (1997) Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res* 57: 1222–1227.
25. Tischkowitz M, Xia B, Sabbaghian N, Reis-Filho JS, Hamel N, et al. (2007) Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci U S A* 104: 6788–6793.
26. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104: 20007–20012.
27. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, et al. (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434: 913–917.
28. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, et al. (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434: 917–921.

### 3.3. Recherche d'un gène siégeant dans les régions délétées des chromosomes 13 et 14 dont l'inactivation serait nécessaire à la carcinogénèse mammaire liée à BRCA2

La délétion récurrente des bras longs des chromosomes 13 et 14 suggère la présence au niveau des régions communes de délétion d'un facteur génétique susceptible de s'associer à la cancérisation mammaire liée à BRCA2. L'inactivation bi-allélique somatique de ce facteur génétique, par, d'une part la délétion d'un allèle, et d'autre part la mutation ponctuelle du second allèle favoriserait la prolifération des tumeurs BRCA2. Afin d'étayer notre hypothèse, une recherche de mutations ponctuelles par une stratégie de séquençage nouvelle génération a été entreprise en réalisant le séquençage de l'exome des bras longs des chromosomes 13 et 14 sur une série de 5 cancers du sein liés à *BRCA2* et l'ADN constitutionnel associé. Ainsi, à l'aide du NGS, nous souhaitons caractériser des altérations associées aux tumeurs BRCA2.

#### 3.3.1. Présentation de la technique NGS

##### 3.3.1.1. De la préparation des échantillons au séquençage

Le séquençage et l'analyse bioinformatique de 5 tumeurs BRCA2 (Tumeurs 52, 106, 133, 148 et 149) et 5 ADN constitutionnels associés ont été réalisés par la société Integragen sur un HiSEQ2000 Illumina au niveau des régions communes de délétion des chromosomes 13 et 14 élargies jusqu'à l'annotation des variants.

L'ADN génomique (3µg) a été fragmenté par sonication et purifié afin d'obtenir des fragments de 150 à 200 pb. Des adaptateurs ont été ligés sur ces fragments, puis purifiés et enrichis par 4 à 6 cycles de PCR. Ces librairies (500ng) ont ensuite été hybridées sur la banque SureSelect (Agilent), puis éluées et amplifiées par 10 à 12 cycles de PCR. Le séquençage sur HiSEQ a permis ensuite de lire 76pb à chaque extrémité des fragments (séquençage paired-end). La banque de capture a été dessinée de façon à obtenir des oligos de 120 mers décalés de 60pb, ce qui permet un recouvrement de chaque base de 2X. Cette banque couvre les exons en 5' et 3' à plus ou moins 100 bases et capture l'exome des bras longs des chromosomes 13 et 14 (chromosome 13 : 19,408,492-99,992,829 (hg19) et chromosome 14 : 22,037,883-106,945,018 (hg19)). Le séquençage réalisé est un séquençage paired-end à une

Profondeur moyenne de 500X, c'est-à-dire que chaque base est lue en moyenne 500 fois. Le séquençage est basé sur l'amplification, accrochage et liaison sur une surface solide dans une *flow cell* qui contient 8 *lanes* ou canaux et l'utilisation de terminateurs de chaînes réversibles marqués par un fluorochrome. Le séquenceur traduit ensuite ces signaux en séquences (« *base calling* »).

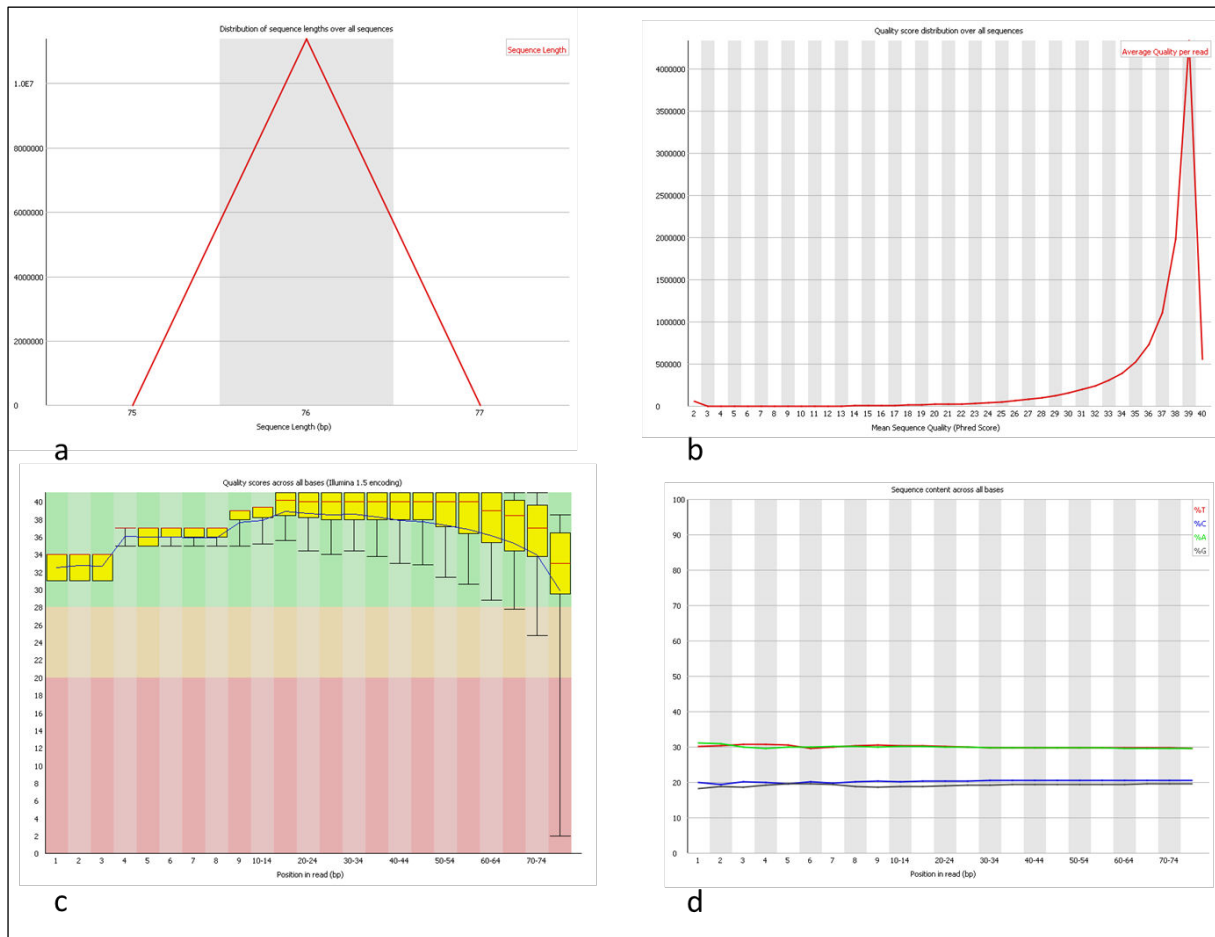
### 3.3.1.2. L'analyse bioinformatique

#### a. Différentes étapes de l'analyse bioinformatique

Pour chaque échantillon, un fichier au format .fastq est généré (2 fichiers .fastq car le séquençage est paired-end). La qualité de ces séquences peut être vérifiée à l'aide du logiciel FastQC [191] (Figure 9) selon les 11 rapports suivants :

- le rapport « Basics Statistics » qui rend les caractéristiques des séquences (codage ASCII, nombre total de séquences...)
- le rapport « Per Base Sequence Quality » évalue le score de qualité moyen des bases en fonction de leur position sur les séquences (le plus bas quartile du QPhred doit être supérieur à 5 et la médiane supérieure à 20)
- le rapport « Per Sequence Quality » évalue le score de qualité Phred moyen des séquences (le QPhred doit être majoritairement >20)
- le rapport « Per Base Sequence Content » calcule le pourcentage des bases A, T, G et C dans les séquences (il ne doit pas y avoir plus de 20% de différences dans une librairie randomisée)
- le rapport « Per Base GC Content » donne le pourcentage de GC pour chaque position (ce pourcentage ne doit pas varier de plus de 10% de la moyenne)
- le rapport « Per Sequence GC Content » donne le pourcentage de GC pour chaque séquence (ce pourcentage doit avoir une distribution normale dans plus de 70% des séquences)
- le rapport « Per Base N Content » reporte le pourcentage de bases non déterminées annotées « N » en fonction de leur position sur la séquence (ce pourcentage doit être inférieur à 20)
- le rapport « Sequence Length Distribution » affiche la taille des séquences
- le rapport « Duplicate Sequences » donne le degré de duplication des séquences qui peut être lié à un biais d'enrichissement (le taux de séquences dupliquées ne doit pas dépasser 50%)
- le rapport « Overrepresented Sequences » signale les séquences présentes plusieurs fois qui peuvent soit expliquer une conséquence biologique importante ou une contamination (une séquence ne doit pas représenter plus de 1% des séquences totales)

-le rapport « Overrepresented Kmers » identifie les séquences de 5 mers sur-représentées en fonction de leur position sur les séquences.



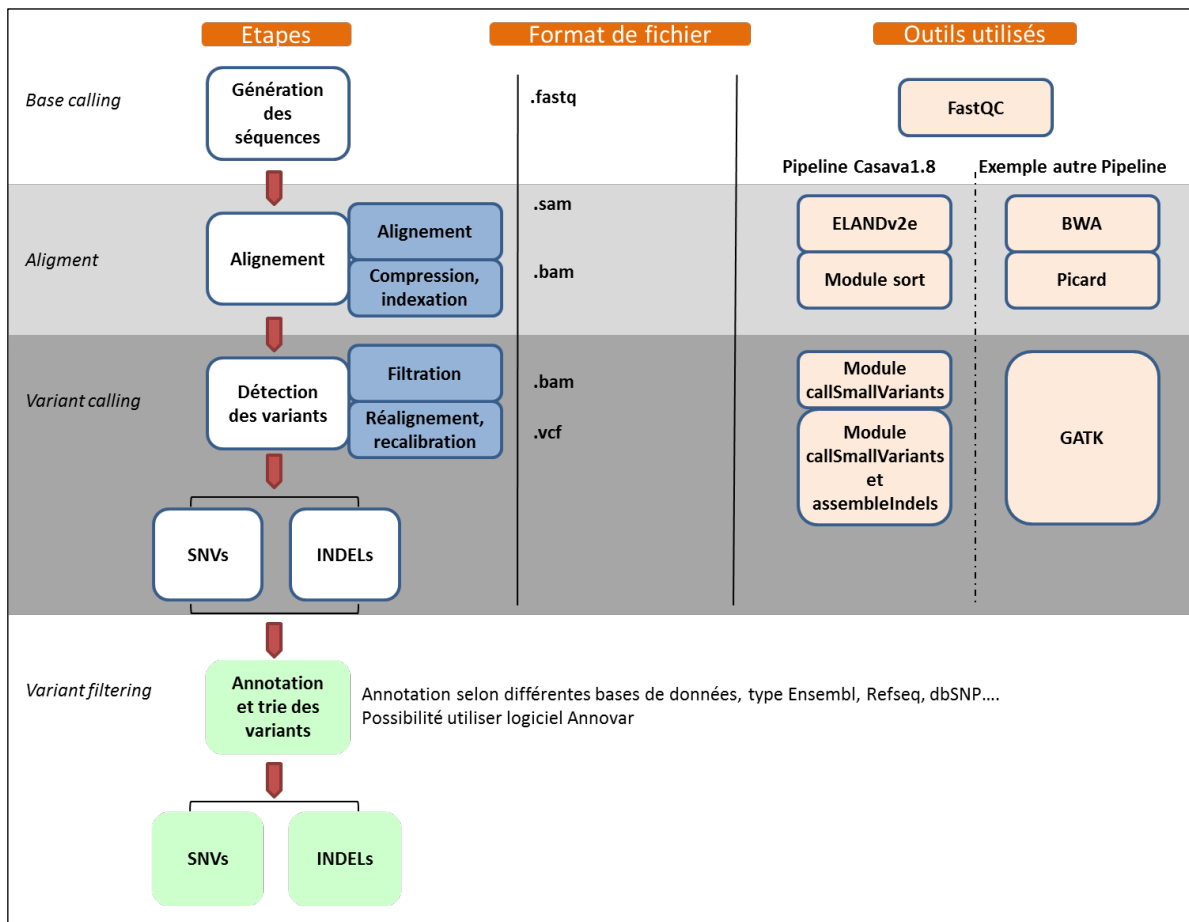
**Figure 9 : Extrait de l'analyse des données fastq par le logiciel FastQC [191]. Exemple de la tumeur 133 (a) Représentation graphique de la taille des séquences, le pic est situé à 76pb. (b) Représentation graphique du score de qualité Phred moyen des séquences, la moyenne est de  $Q_{Phred}=39$ , ce qui signifie une probabilité d'erreur inférieure à  $10^{-4}$ . (c) Représentation graphique des scores de qualité moyens des bases en fonction de leur position sur les reads. On observe une diminution de qualité importante à partir de la base 60 ( $Q_{Phred}<20$ ). (d) Représentation graphique du pourcentage de GC au sein des séquences, ici 40% GC et 60% AT.**

L'analyse bioinformatique des analyses NGS se résume 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. L'analyse des données de séquençage fournie par Integragen est basée sur le pipeline Illumina CASAVA 1.8, mais il existe d'autres pipelines tels que l'association de BWA (*Burrows Wheeler Alignment Tool*), Picard (<http://picard.sourceforge.net/>) et GATK (*Genome Analysis Toolkit*) [192;193] (Figure 10).

À partir des données .fastq, les séquences ou *reads* sont alignées sur le génome de référence et converties en .sam (*Sequence Alignment/Map*). Pour l'alignement, différents algorithmes tels que ELANDv2e et BWA sont disponibles. Les fichiers .sam sont ensuite compressés, indexés et convertis

au format binaire .bam. Les *reads* dupliquées sont éliminées. Cette étape est réalisée par le module « sort » de Casava, et peut également être réalisée à l'aide du logiciel Picard. Les fichiers .bam subissent un nouveau réalignement car les INDELS (pour Insertion-Délétion) peuvent créer de faux SNVs (pour *Single Nucleotide Variation*), et subissent également une recalibration des scores de qualité des bases par les modules « callSmallVariants » du *pipeline* Casava ou par l'algorithme GATK. Des fichiers .vcf (*variant call format*) sont ensuite générés par les modules « callSmallVariants » (revoit les SNVs et petites INDELS) et « assembleIndels » (revoit les grandes INDELS) pour le *pipeline* Casava ou par GATK. Les fichiers .vcf revoit les différentes variations SNVs ou INDELS détectées par rapport au génome de référence.

Les fichiers .vcf bruts ne sont pas annotés, cela signifie que les variations relevées sont exprimées en position génomique, allèle de référence/allèle variant et des données de qualité comme la profondeur de lecture et le score de qualité du variant. La dernière étape consiste alors à annoter ces fichiers à l'aide de bases de données telles que Ensembl, Refseq pour connaître le gène, la position exonique ou intronique du variant, la conséquence protéique de la mutation ; mais aussi des bases de données telles que dbSNP, 1000Genome qui donneront des indications sur la fréquence de la variation. Il est possible d'utiliser d'autres bases de données et logiciels comme l'algorithme de prédiction PolyPhen par exemple. L'annotation peut s'effectuer avec le logiciel Annovar. 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 a été généré.



**Figure 10 : Résumé des différentes étapes bioinformatiques à partir de données de séquençage haut-débit. Le pipeline Casava1.8 utilisé dans l'étude ainsi que l'exemple d'un autre pipeline sont détaillés.**

*b. Description des formats de fichiers*

Le format .fastq

Le format fastq est un format de fichier texte permettant de stocker des séquences nucléiques associées à des scores de qualité. Par analogie avec le format FASTA, le « < » est remplacé par « @ ». Le score de qualité en base de 10 appelé QPhred de la séquence est codé avec un caractère ASCII qui est une norme de codage de caractères en informatique (Figure 11).

Le QPhred correspond à une probabilité  $p$  d'erreur de mauvaise identification de la base et est donné par la formule suivante :  $QPhred = -10 \log_{10} p$ . Ainsi, un QPhred de 30 correspond à une probabilité d'erreur de 0,1%. L'échelle des scores de qualité Phred et le codage en caractère ASCII diffèrent selon les plateformes de séquençage. Le décalage entre le score de qualité 0 et le caractère ASCII correspondant est symbolisé par « Phred+33 » ou « Phred+64 ».





## Le format .sam

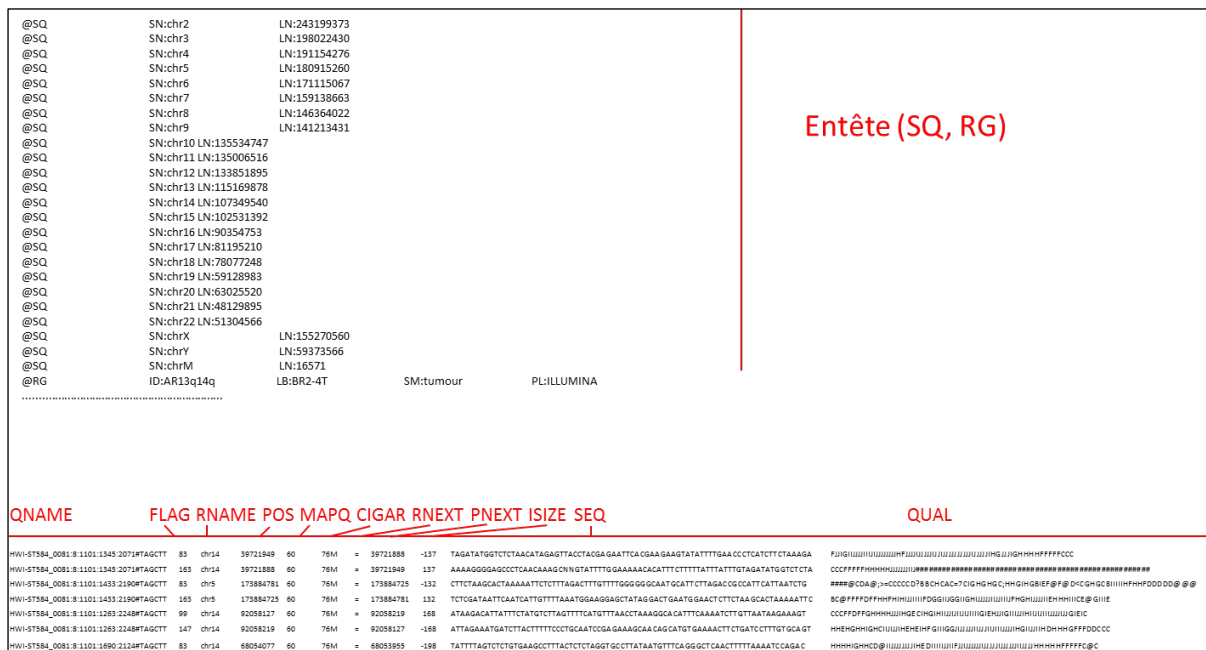
Le format .sam pour *Sequence Alignment/Map* format est un fichier au format texte TAB-delimited constitué d'une entête suivie de données sur les séquences et l'alignement [195]. Pour différencier les deux parties, dans l'entête chaque ligne commence par le symbole @ (Figure 13).

L'entête peut rendre diverses informations telles que le dictionnaire des séquences de référence symbolisé par « SQ » et le *read group* (RG). Un même RG est assigné à des *reads* issus du même *run* de séquençage.

En-dessous de l'entête, les données s'articulent en 11 colonnes obligatoires qui renseignent des informations sur l'alignement pour chaque *read* (Table 9). Les valeurs de ces colonnes peuvent être « 0 » ou « \* » si aucune information n'est donnée. Il est possible d'ajouter à ces 11 champs obligatoires, des données optionnelles qui peuvent également être utiles pour juger de la qualité des *reads*. Pour exemple, les champs optionnels peuvent rapporter le nombre de fois où le *reads* peut être aligné sur la séquence de référence.

No.	Nom	Description
1	QNAME	Nom du <i>read</i> (nom identique au fichier .fastq)
2	FLAG	Annotation de l'alignement du <i>read</i> en codage binaire (non aligné, alignement sur brin sens ou antisens, orientation des <i>reads</i> d'un séquençage paired-end...)
3	RNAME	Nom de la séquence de référence
4	POS	Position du début de l'alignement du <i>read</i> dans un système de coordonnées où la première base d'une séquence est 1
5	MAPQ	Qualité d'alignement (score Phred de la probabilité d'un mauvais alignement)
6	CIGAR	Représentation d'un alignement avec des codes simples (ex : M = match ou mismatch, I = insertion, D = délétion...).
7	RNEXT	Nom de la séquence de référence du <i>read</i> partenaire
8	PNEXT	Position du <i>read</i> partenaire
9	ISIZE	Longueur de l'insertion entre les deux <i>reads</i> partenaires
10	SEQ	Séquence du <i>read</i>
11	QUAL	Score Qphred de chaque base de la séquence (identique au score Qphred du fichier fastq)

**Table 9: Présentation des 11 champs obligatoires du format .sam *Sequence Alignment/Map* en séquençage paired-end [195].** Pour exemple, un CIGAR de 3I4M1D indique une insertion de 3 bases suivie de 4 bases alignées et 1 base délétée.



**Figure 13 : Extrait d'un fichier .sam *Sequence Alignment/Map*, exemple de la tumeur 52. L'entête contient particulièrement le dictionnaire des séquences de référence « SQ » et le read group « RG ». Chaque read est défini par une ligne renseignant 11 champs obligatoires qui reprennent des informations sur l'alignement : QNAME, FLAG, RNAME, POS, MAPQ, CIGAR, RNEXT, PNEXT, ISIZE, SEQ, QUAL.**

### Le format .bam

Le format .bam est une compression et indexation du format .sam (algorithme BGZF) et constitue une version binarisée de .sam [195]. Un fichier .bam est associé à un fichier d'index .bai, ce qui permet de faciliter l'accès à une position dans le fichier bam. Les données .bam peuvent être visualisées à l'aide de logiciels tels que IGV (*Integrative Genomics Viewer*).

### Le format .vcf

Le format .vcf (*variant call format*) a été proposé par Danecek *et al.*, pour standardiser le format des fichiers de variants issus de l'analyse NGS [196].

Un fichier .vcf est constitué d'une partie de titre et d'une partie de données. Le titre contient des informations sur les annotations et les tags utilisés dans la partie données, mais peut également contenir d'autres informations telles que la date de création, les logiciels utilisés. Chaque ligne commence par le symbole « ## » sauf une ligne de titre définissant les 8 colonnes de la partie données qui commence par le symbole « # ». Les 8 colonnes de données fournissent les informations suivantes: le chromosome (CHROM), la position du variant (POS), le nom du variant (ID), l'allèle de

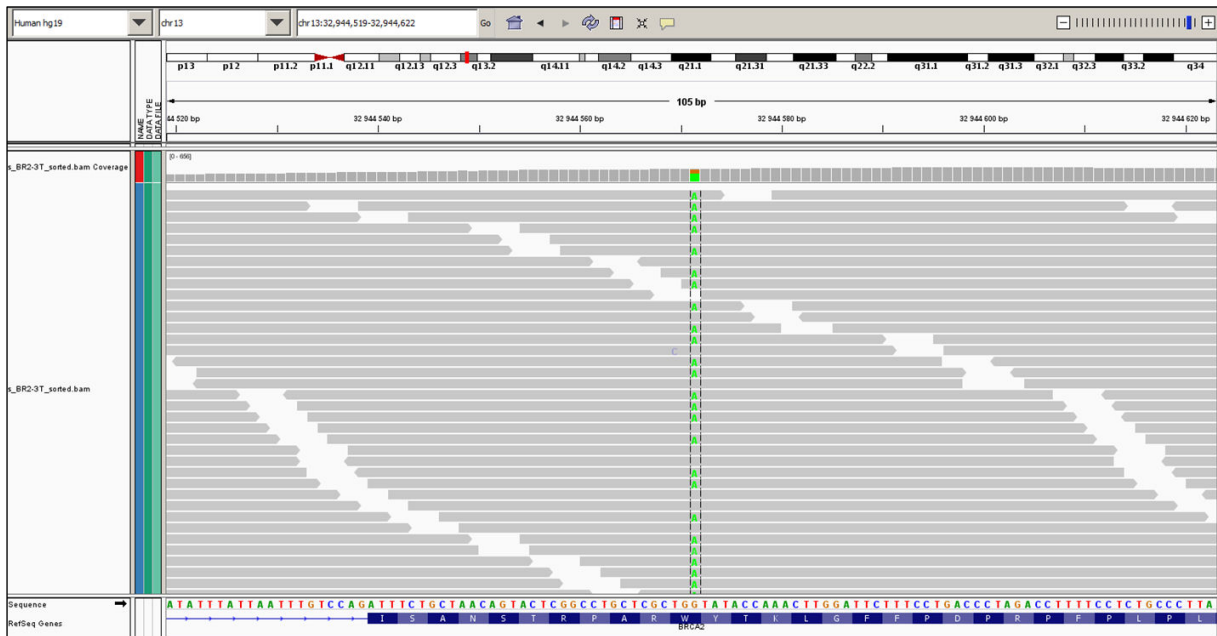
référence (REF), la liste des allèles alternatifs séparés par des virgules (ALT), un score de qualité en Phred (QUAL), le statut du variant selon les critères définis dans la partie titre (FILTER), et une dernière colonne d'annotation (INFO) contenant des codes clés optionnels séparés par des points virgules informant entre autre sur l'allèle ancestral, le CIGAR et la profondeur (Figure 14).

```
##Header lines
##fileformat=VCFv4.1
.....
#CHROM POS ID REF ALT QUAL FILTER INFO
chr13 49791536 rs2181906 A C 12.77 LowQual ABHom=1.00;AC=2;AF=1.00;AN=2;DB;DP=1;Dels=0.00;FS=0.000;HaplotypeScore=0.0000;MLEAC=2;MLEAF=1.00;MQ=60.00;MQ0=0;QD=12.77
chr13 49823287 rs3764131 A T 850.77 PASS ABHom=1.00;AC=2;AF=1.00;AN=2;DB;DP=24;Dels=0.00;FS=0.000;HaplotypeScore=0.0000;MLEAC=2;MLEAF=1.00;MQ=59.03;MQ0=0;QD=23.82
chr13 49832383 rs9591249 T C 12.77 LowQual ABHom=1.00;AC=2;AF=1.00;AN=2;DB;DP=1;Dels=0.00;FS=0.000;HaplotypeScore=0.0000;MLEAC=2;MLEAF=1.00;MQ=60.00;MQ0=0;QD=12.77
chr13 49841499 rs7334929 T G 2111.77 PASS ABHom=1.00;AC=2;AF=1.00;AN=2;DB;DP=58;Dels=0.00;FS=0.000;HaplotypeScore=0.0000;MLEAC=2;MLEAF=1.00;MQ=59.60;MQ0=0;QD=28.08
chr13 49848711 rs7490591 A G 551.77 PASS ABHom=1.000;AC=2;AF=1.00;AN=2;DB;DP=17;Dels=0.00;FS=0.000;HaplotypeScore=0.0000;MLEAC=2;MLEAF=1.00;MQ=38.90;MQ0=0;QD=32.46
```

**Figure 14 : Extrait d'un fichier .vcf variant call format, exemple de la tumeur 52.**

### c. Visualisation

La visualisation des variants peut se faire à l'aide d'IGV téléchargeable à l'adresse suivante <http://www.broadinstitute.org/igv/> (Figure 15). Cette représentation permet d'observer l'alignement de chaque *read* sur la séquence de référence et d'observer les variations nucléotidiques (substitution ou insertion-délétion) présentes dans les *reads* par rapport à la séquence de référence (ici par rapport à la version hg19 du génome humain). Ainsi, cette visualisation des *reads* permet d'accéder directement le nombre de *reads* à une position génomique donnée et le nombre de *reads* représentant la variation nucléotidique (plusieurs exemples sont donnés en Annexe 3).



**Figure 15 :** Visualisation par IGV d'un fichier .bam, exemple de la tumeur 52. Substitution nucléotidique G>A en position génomique chr13 :32.944.571 (hg19) de l'exon 19 du gène BRCA2.

### 3.3.2. Stratégie d'analyse des données NGS

Le but de l'analyse NGS est de rechercher des variants somatiques ciblant un gène commun aux 5 tumeurs BRCA2, afin d'identifier un événement somatique potentiel récurrent. Pour cela, plusieurs démarches d'analyse ont été réalisées.

#### 3.3.2.1. Description des fichiers résultats obtenus

L'analyse bioinformatique par la société Integragen (*pipeline* CASAVA1.8) des données de séquençage haut débit de 5 tumeurs BRCA2 et de leur ADN constitutionnel a abouti à la génération de fichiers de variants annotés par échantillon : un fichier de variants SNVs et un fichier de variants INDELS pour chaque tumeur et un fichier de variants SNVs et un fichier de variants INDELS pour les échantillons constitutionnels. Soit un total de 5 fichiers SNVs (tumeur), 5 fichiers SNPs (constitutionnel) et 10 fichiers INDELS (5 fichiers INDELS tumeur et 5 fichiers INDELS constitutionnel).

Ces variants sont ordonnés selon la position génomique du variant sur le génome humain version hg19. Ainsi, un variant peut être répertorié plusieurs fois s'il existe plusieurs transcrits de référence. Toutes les manipulations de fichier de type « comparaison » ou « tri » ont été codées en langage R (script maison de comparaison des fichiers .txt des variants SNPs, SNVs et INDELS) afin d'avoir une analyse automatisée et reproductible.

En plus des fichiers de variants, un fichier comparant les variants détectés dans la tumeur aux variants de l'ADN constitutionnel correspondant dénommé « variants selon statut » a été généré par une analyse statistique. Cette analyse permet d'attribuer pour chaque couple d'échantillon tumeur-constitutionnel le statut somatique ou constitutionnel de chaque variant. La détermination du statut somatique ou constitutionnel du variant a été réalisée à l'aide du module « Germline » de CASAVA1.8. Les positions détectées comme étant variantes selon la séquence de référence sont soumises à un test statistique de Fisher (fonction `fisher.test` du *package* `stats` du logiciel R). Ce test calcule s'il existe ou non une divergence significative entre les fréquences alléliques du couple Constit/Tumeur. Enfin, après une comparaison des génotypes entre les échantillons constitutionnels et tumoraux réalisés, un statut « constitutionnel », « constitutionnel potentiel », « somatique », « somatique potentiel », « perte d'hétérozygotie » ou « non calculé » est attribué à chaque variant. Le statut non calculé concerne les variants dont la profondeur de lecture est inférieure à 8 chez l'un ou l'autre des échantillons ou dont le ratio (nombre de *reads* éliminées/nombre de *reads* utilisées) est supérieur à 0.5.

### 3.3.2.2. Analyse des variants annotés

#### a. Recherche de variants SNVs

Nous avons tout d'abord étudié le fichier « variants selon statut » pour isoler les variants annotés somatiques par la méthode statistique décrite précédemment. Nous avons sélectionné les 49 variants annotés « somatique » puis les avons triés selon leur position exonique (15 SNVs) ou intronique (34 SNVs) pour ne garder uniquement que les 15 variants « somatiques » exoniques. Treize SNVs « somatiques » exoniques correspondent à des mutations de type faux-sens, un SNV est une mutation de type non-sens, et le dernier est une mutation iso-sémantique (Table 10a). Nous avons choisi de ne pas retenir les variants iso-sémantiques à l'exception de ceux positionnés dans les sites donneurs consensus exoniques (GT 5')/accepteur (AG 3') d'épissage.

Au total, nous avons obtenu 14 SNVs « somatiques » exoniques dont 13 faux-sens et 1 non-sens.

Nous avons également étudié les variants « somatiques potentiels ». Ces variants partagent tous la même caractéristique qui est une forte variation du ratio allélique entre la tumeur et le constitutionnel. Ce sont des SNPs présents à l'état hétérozygote dans le constitutionnel, mais qui sont la cible de la perte d'un allèle ou LOH dans la tumeur. Un exemple est donné en Annexe 3d. Or, nous souhaitons caractériser un évènement mutationnel somatique qui inactiverait le second allèle d'un gène par la survenue d'une mutation ponctuelle dans la tumeur et non la perte d'un allèle. L'autre allèle du gène étant perdu car localisé sur la région minimale de délétion des chromosomes 13 et 14, caractéristique partagée par les tumeurs BRCA2 étudiées. Nous ne poursuivons pas l'étude des variants « somatiques potentiels ». Il y a cependant une exception sur les 446 variants « somatiques potentiels » : le variant faux-sens G>A sur le gène *PARP4* en position génomique 25.029.218, rs2275660 caractérisé dans la tumeur 106. L'allèle mineur est présent à 98% et l'allèle de référence à 2% mais l'ADN constitutionnel est homozygote pour l'allèle mineur. Soit la présence de l'allèle de référence à 2% dans la tumeur signifie l'acquisition d'un clone avec l'allèle de référence, soit il s'agit d'un faux positif. Ce SNV ne constitue en aucun cas un variant candidat pour notre hypothèse, car la variation nucléotidique entre la tumeur et le constitutionnel conduit à l'acquisition de l'allèle de référence dans la tumeur.

Nous avons repris l'analyse par une autre méthode pour isoler à nouveau les variants « somatiques » de type SNV. Pour cela, nous avons comparé les fichiers de variants SNVs de la tumeur et du constitutionnel alignés sur la position génomique et le génotype. Nous avons poursuivi l'analyse avec les 861 variants présents uniquement dans la tumeur (Table 10b), ce qui représente 3 variants non-sens, 113 faux-sens, 91 variants iso-sémantiques dont 3 positionnés dans les sites donneurs (GT 5')/ accepteurs (AG 3') d'épissage, et 654 variants introniques dont 5 à +/- 4 nucléotides de l'exon pouvant avoir un retentissement sur l'épissage. Comme précédemment, nous avons choisi de ne pas retenir les variants iso-sémantiques s'ils ne sont pas positionnés dans les sites donneurs (GT 5')/ accepteurs (AG 3') d'épissage ni les variants introniques hormis ceux situés à +/- 4 nucléotides de l'exon pouvant avoir un retentissement sur l'épissage.

Avec cette démarche, nous avons donc isolé 124 SNVs somatiques, et plus précisément, 3 variants non-sens, 113 faux-sens, 3 variants iso-sémantiques positionnés dans les sites donneurs (GT 5')/ accepteurs (AG 3') d'épissage, et 5 variants introniques à +/- 4 nucléotides de l'exon.

Tumeur	Méthode statistique: SNVs "somatiques" exoniques			SNVs "somatiques" introniques		
	non-sens	faux-sens	iso-sémantique +/- 2 NT	iso-sémantique	intronique +/- 4 NT	intronique
52	0	1	0	1	0	1
106	0	0	0	0	0	8
133	0	0	0	0	0	2
148	0	5	0	0	0	12
149	1	7	0	0	0	11
<b>Total</b>	1	13	0	1	0	34

a

Tumeur	Méthode non statistique: SNVs "somatiques" exoniques			SNVs "somatiques" introniques		
	non-sens	faux-sens	iso-sémantique +/- 2 NT	iso-sémantique	intronique +/- 4 NT	intronique
52	0	11	1	13	0	68
106	0	14	0	4	0	65
133	2	6	1	7	0	37
148	0	65	0	56	5	408
149	1	17	1	8	0	71
<b>Total</b>	3	113	3	88	5	649

b

**Table 10: Description des SNVs somatiques observés dans les tumeurs BRCA2 séquencées en NGS.** (a) SNVs somatiques déterminés par un test t de Fisher, (b) SNVs somatiques déterminés par la comparaison des fichiers de variants caractérisés dans l'ADN tumoral et l'ADN constitutionnel apparié. Les variants « iso-sémantiques +/- 2 NT » sont des variants positionnés dans les sites donneurs (GT 5')/accepteur (AG 3') d'épissage et les variants « intronique +/- 4NT » sont des variants introniques positionnés à +/- 4 nucléotides de l'exon. Ces variants peuvent avoir un retentissement sur l'épissage.

Nous avons ensuite confronté les résultats obtenus par la première méthode statistique et la seconde méthode de comparaison non statistique permettant d'isoler les variants somatiques.

Les 14 SNVs annotés « somatiques » par un test statistique ont été retrouvés par la méthode non statistique. Nous avons visualisé les fichiers .bam sur IGV pour observer les reads de l'échantillon tumoral et de l'ADN constitutionnel associé aux positions génomiques des 14 SNVs « somatiques ». Pour 4 SNVs « somatiques », la visualisation des données .bam a montré la présence des variants nucléotidiques à la fois dans l'échantillon tumoral et le constitutionnel associé alors que le statut de ces variants a été déterminé « somatique » par l'algorithme « Germline » de Casava. En examinant les données utilisées pour la détermination du statut « constitutionnel » ou « somatique » des variants qui donnent accès aux nombres de bases A, T, G et C séquencées à une position génomique, nous avons observé que pour ces 4 SNVs, les variations nucléotidiques sont présentes dans l'ADN tumoral et constitutionnel. Ces variants ont été annotés « somatiques » à tort car l'algorithme utilisé a calculé un génotype homozygote pour l'allèle de référence dans l'échantillon constitutionnel. Il est possible que les reads contenant le variant dans l'échantillon constitutionnel aient été supprimées pour l'analyse en raison d'une mauvaise qualité. Trois de ces SNVs sont localisés sur le gène *PABPC3* de la tumeur 149, dont un variant non-sens, et un SNV est localisé sur le gène *AHNAK2*. Le séquençage en Sanger



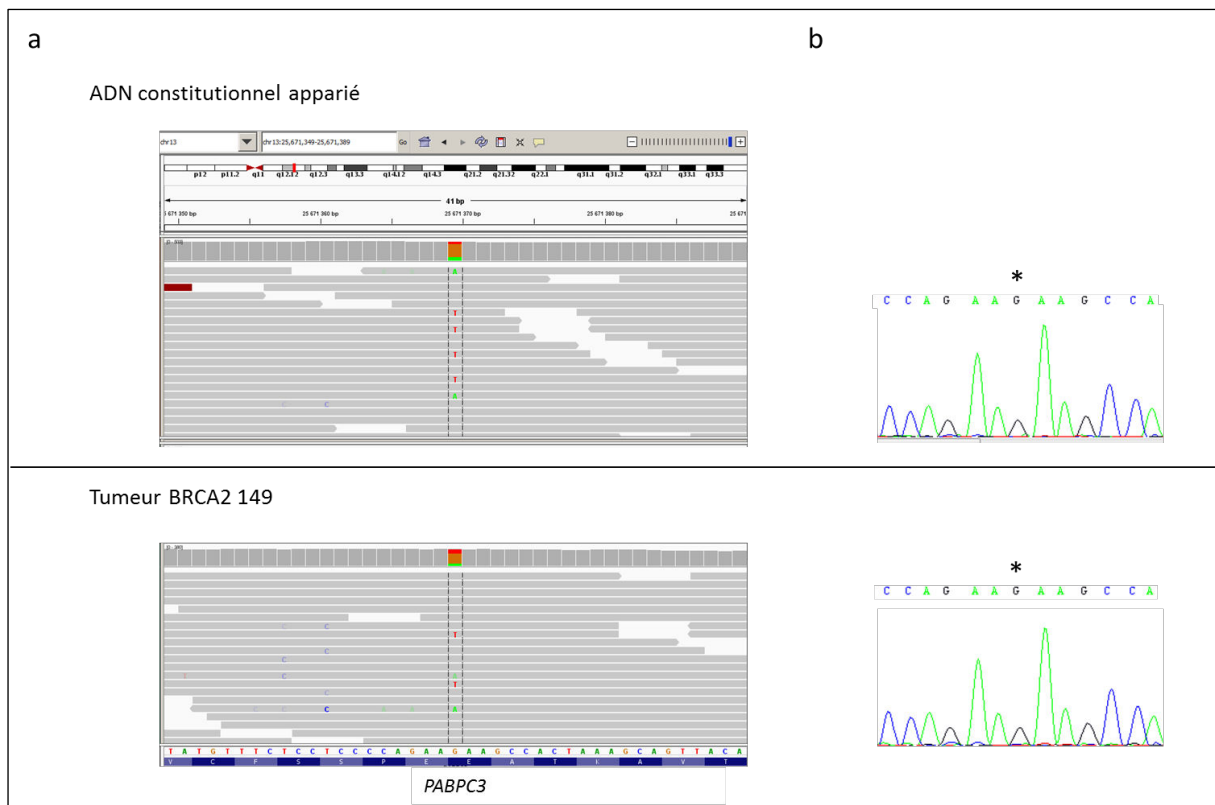
n'a pas permis de caractériser le variant non-sens sur l'ADN tumoral et l'ADN constitutionnel a priori (Figure 16). De plus, en soumettant ces séquences à l'algorithme BLAT en vue de caractériser des régions homologues à ces séquences [197], nous observons que ces gènes présentent de fortes régions d'homologie avec d'autres gènes. De plus le ratio allélique observé sur le constitutionnel est d'environ 30/70 alors qu'il doit être proche de 50/50. Ces constatations amènent à penser que ces 4 SNVs sont des faux-positifs provenant d'un mauvais alignement et ont conduit à l'exclusion de ces 4 SNVs.

Pour conclure, la combinaison des deux analyses a révélé 10 SNVs « somatiques » localisés sur 10 gènes différents dont 8 sur le chromosome 14 et 2 sur le chromosome 13. Nous avons relevé 5 variants faux-sens dans la tumeur 148 localisés sur les gènes *KATNAL1*, *RABGGTA*, *CTAGE5* et *FAM178B*, et 5 variants faux-sens dans la tumeur 149 localisés sur les gènes *ELF1*, *FANCM*, *BMP4*, *UNC79* et *TRAF3* (Table 11). Aucun SNV somatique exonique n'a été caractérisé sur les tumeurs 52, 106 et 133 et chaque variant n'est présent que dans une tumeur.

Sur les 110 autres SNVs somatiques exoniques détectés par la méthode non statistique, 33 SNVs (2 non-sens et 31 faux-sens) ont une profondeur de détection inférieure à 8, seuil utilisé par le pipeline Casava. Nous avons donc exclu les 31 faux-sens mais avons vérifié par un séquençage Sanger le statut somatique des 2 variants non-sens (gène *SPERT* et *TTC9*). Le variant du gène *SPERT* est présent dans la tumeur et le constitutionnel a priori (Annexe 3b). La fréquence dans la population générale du variant non-sens du gène *SPERT* rs79707842 est rare selon les données du projet NGS NHLBI *Exome Sequencing Project* (ESP), et estimée à 0.0549 dans la population caucasienne [198]. Il s'agit donc d'un polymorphisme considéré à tort comme somatique du fait d'une faible profondeur de lecture. Le variant non-sens du gène *TTC9* n'a pas été retrouvé ni dans la tumeur, ni dans le constitutionnel a priori par séquençage Sanger. Les 77 variants restants (5 variants introniques à +/- 4 nucléotides de l'exon, 68 variants faux-sens, et 3 variants iso-sémantiques positionnés dans les sites donneurs ou accepteurs d'épissage) ont été visualisés dans IGV mais sur les fichiers .bam, ces variants sont présents à la fois sur le constitutionnel et la tumeur, ils n'ont donc pas été retenus (faux positif en raison d'un mauvais alignement ou polymorphisme constitutionnel). Pour ces raisons, nous ne poursuivrons l'analyse pour aucun des 110 SNVs somatiques exoniques détectés uniquement par la méthode non statistique. Cette méthode n'a pas permis de caractériser de SNVs somatiques exoniques d'intérêt supplémentaires.

Tumeur	Chromosome	Position hg19	Gène	Allèle de référence	A	C	G	T	Profondeur	Q.snp.	Conséquence
148	13	30782807	KATNAL1	G	194	0	368	0	562	770	c.1343C>T;p.Ser448Phe
149	13	41515163	ELF1	G	72	0	119	0	191	338	c.1150C>T;p.Arg384Trp
148	14	24737356	RABGGTA	T	9	0	0	26	35	85	c.1107A>T;p.Glu369Asp
148	14	39818131	CTAGE5	A	341	0	0	207	548	800	c.2198A>T;p.Asp733Val
148	14	45433620	FAM179B	C	106	216	0	0	322	394	c.1996C>A;p.Pro666Thr
149	14	45669143	FANCM	C	0	106	0	71	177	424	c.6079C>T;p.His2027Tyr
148	14	52508864	NID2	G	0	0	47	26	73	160	c.1784C>A;p.Ala595Asp
149	14	54417259	BMP4	G	32	0	65	0	97	152	c.718C>T;p.His240Tyr
149	14	94110036	UNC79	G	127	0	152	0	279	757	c.5623G>A;p.Val1875Met
149	14	103369664	TRAF3	G	0	11	28	0	39	42	c.1033G>C;p.Glu345Gln

**Table 11 : Description des 10 SNVs somatiques étudiés en Sanger observés dans les tumeurs BRCA2 séquencées en NGS.** Les colonnes A, C, G et T, renseignent sur le nombre de lectures des 4 bases à une position génomique donnée, la profondeur de détection qui renseigne sur le nombre de séquences est donnée par la Profondeur, le Q.snp. est le score de qualité Phred du variant (probabilité que le SNP soit une erreur).



**Figure 16 : Etude du variant non-sens c.1033G>T ; p.Glu345X du gène PABPC3 sur la tumeur 149 et l'ADN constitutionnel apparié.** (a) La visualisation des fichiers .bam a montre la présence de la variation nucléotidique c.1033G>T et c.1033G>A à la fois dans l'échantillon tumoral et le constitutionnel apparié, le ratio allélique dans la tumeur et le constitutionnel est d'environ 20% nucléotide A, 20% nucléotide T, 60% nucléotide G. (b) Le séquençage Sanger n'a pas caractérisé le variant non-sens sur l'ADN tumoral et l'ADN constitutionnel apparié.

b. Recherche de variants INDELS

Pour la caractérisation des variants INDELS, la société Integragen n'a fourni que les fichiers de variants sans réaliser l'analyse statistique permettant d'attribuer pour chaque couple d'échantillon tumeur-constitutionnel le statut somatique ou constitutionnel de chaque variant. Nous avons donc isolé les variants somatiques INDELS par la méthode non statistique décrite précédemment en comparant les fichiers de variants INDELS de la tumeur et du constitutionnel attribué sur la position génomique. Nous avons poursuivi l'analyse avec les 186 variants INDELS présents uniquement dans la tumeur, ce qui représente 19 variants exoniques, 188 variants introniques dont 2 situés à +/- 4 nucléotides de l'exon pouvant avoir un retentissement sur l'épissage (Table 12).

Pour la suite de l'analyse, nous allons étudier les 19 INDELS exoniques somatiques et les 2 INDELS introniques situés à +/- 4 nucléotides de l'exon. Comme précédemment pour la caractérisation des SNVs somatiques, nous avons visualisé les fichiers d'alignement .bam sur IGV. Nous avons observé la présence de l'INDEL sur l'échantillon tumoral et le constitutionnel attribué pour 8 variants qui seront exclus.

Pour conclure, nous avons obtenu 6 INDELS exoniques localisées uniquement sur le chromosome 14, sur 6 gènes différents, dans les tumeurs 106, 133, 148 et 149 dont une délétion qui n'est pas en phase et qui entraîne l'arrivage d'un codon stop prématuré (gène *PRMT5*). Un variant INDEL est particulier car son insertion ne peut être bornée, une portion de la séquence insérée est portée dans la Table 13 (gène *PCNX*). Chaque variant n'est présent que dans une tumeur, et aucun n'est présent dans la tumeur 52 (Table 13).

Tumeur	INDELS "somatiques" exoniques		INDELS "somatiques" introniques	
	exonique	intronique +/- 4NT	intronique	
52	4	1	33	
106	5	0	32	
133	2	0	35	
148	5	1	51	
149	3	0	35	
<b>Total</b>	<b>19</b>	<b>2</b>	<b>186</b>	

**Table 12 : Description des INDELS somatiques observées dans les tumeurs BRCA2 séquencées en NGS. INDELS somatiques retrouvées exclusivement dans les tumeurs d'après une comparaison des fichiers de variants des couples Constit/Tumeur. Les variants « +/- 4 NT » sont des variants situés à +/- 4 bases du site d'épissage de l'exon.**

Tumeur	Chromosome	Position hg19	Gène	CIGAR	Séquence INDEL	Profondeur	Q.indel.	Lecture INDEL	Conséquence
106	14	23393534	PRMT5	44D	CCTGCTTGGCTGCCCGCA GGGAAGCGTTCACCAGG GGTCCCCGT	250	199	18	c.1101_1144del;p.Gly369ProfsX24
149	14	23856762	MYH6	15D	CAGCTCCAGCTTCTC	101	69	8	c.4612_4626del;p.Glu1538_Leu1542del
106	14	53331283	FERMT2	3I	TGG	225	1371	32	c.1436_1438dup;p.Thr479dup
106	14	53619481	DDHD1	6I	GCCGCC	4	164	3	c.331_336dup;p.Gly111_Gly112dup
133	14	71444986	PCNX	BP_LEFT	TGTCATACTGTCTAGT ATTTTGTTTTTGTTTTT TCCGAGAGAGAGTCTTG CTGTGTCTCCTAGGCTGG AGTGCACTGGCTTGATCT TGGCTCACTGCAAGCTCT GCCT	326	719	29	?
148	14	73732096	PAPLN	30D	CCAGCCCCGGGTGGTGG ATGCCAGTCCAGG	21	319	7	c.3063_3092del;p.Pro1023_Gln1032del

**Table 13 : Description des 6 INDELs somatiques étudiées en Sanger observées dans les tumeurs BRCA2 séquencées en NGS.** La colonne CIGAR indique la nature de l'insertion/délétion. « xD » indique la délétion de x bases, « xI » indique l'insertion de x bases, et « BP\_LEFT » indique que l'insertion /délétion ne peut être bornée. Dans cet exemple, le BP\_LEFT correspond à une insertion d'au moins 112 bases. La colonne séquence INDEL donne la séquence Insérée ou Délétée. La profondeur de détection qui renseigne sur le nombre de séquences est donnée par la « Profondeur », le « Q.indel » est le score de qualité Phred du statut génotypique de l'INDEL (codant la probabilité que le génotype soit faux). Lecture INDEL donne le nombre de séquences incluant l'INDEL.

### 3.3.2.3. Etude des CNA

Les données NGS permettent également l'étude des CNA pour *Copy Number Aberrations*. Pour cela, une analyse à l'aide du package DNACo y (*Bioconductor*) a été réalisée [199]. Ce dernier permet de détecter et localiser les changements brusques dans le nombre de copies d'ADN. La méthode CBS (*Circular Binary Segmentation*) est utilisée pour segmenter les données de séquençage d'Exome [200]. Les données de séquençage de l'ADN tumoral sont comparées aux données de l'ADN constitutionnel. Un segment (une région) de l'ADN tumoral est considéré délété ou amplifié lorsque la valeur du log2 ratio de la profondeur de lecture entre les deux échantillons (c'est-à-dire le tumeur/constit) est supérieure à 0,32 ou inférieure à -0,41, où  $\text{Log2 ratio} = \text{Log2} \left( \frac{\text{profondeur tumeur}}{\text{profondeur constitutionnel}} \right)$ .

Les ratios sont normalisés par le nombre total de séquences et ajustés de façon à ce que la médiane des log2 ratio des exons des chromosomes « normaux » soit de zéro. Dans le fichier résultant de l'analyse, chaque ligne correspond à une région délétée ou gagnée par rapport aux données de séquençage du constitutionnel. La valeur du log2 ratio, du nombre de copies et les coordonnées génomiques sont indiquées. Une représentation graphique sous-forme de plot des log2 ratio en fonction de la position génomique par analogie aux profils CGH et traces SNP est présentée en Figure

17. Il est essentiel de noter la similarité des profils obtenus par analyse des CNA à partir des données NGS et ceux obtenus par analyse génomique en BAC-CGH et puces SNP.

Nous avons analysé ces données pour affiner la zone minimale de délétion des chromosomes 13 et 14 et pour rechercher des cassures intragéniques qui inactiveraient le second allèle d'un gène localisé dans la région minimale de délétion.

a. Etude des régions délétées selon les données NGS

Les coordonnées génomiques des régions perdues selon l'analyse des CNA par le package DNACopy dans les tumeurs BRCA2 sur les chromosomes 13 et 14 ont été relevées. Ainsi, la région minimale de délétion du chromosome 14 selon ces nouvelles données s'étend entre les positions 71.197.422 et 100.381.027. Les bornes de la région minimale de délétion du chromosome 14 d'après l'analyse des profils SNP et CGH est 70.418.189-99.851.630. Pour le chromosome 13, seule la borne télomérique de la délétion en NGS a pu être affinée, car la tumeur 43 qui présentait la délétion centromérique la plus distale n'a pas été étudiée en NGS. Ainsi la borne télomérique de la région minimale de délétion en NGS sur le chromosome 13 est 53.603.006 et avait été évaluée en 53.099.075 d'après l'analyse des profils SNP et CGH.

b. Etude des pertes bi-alléliques

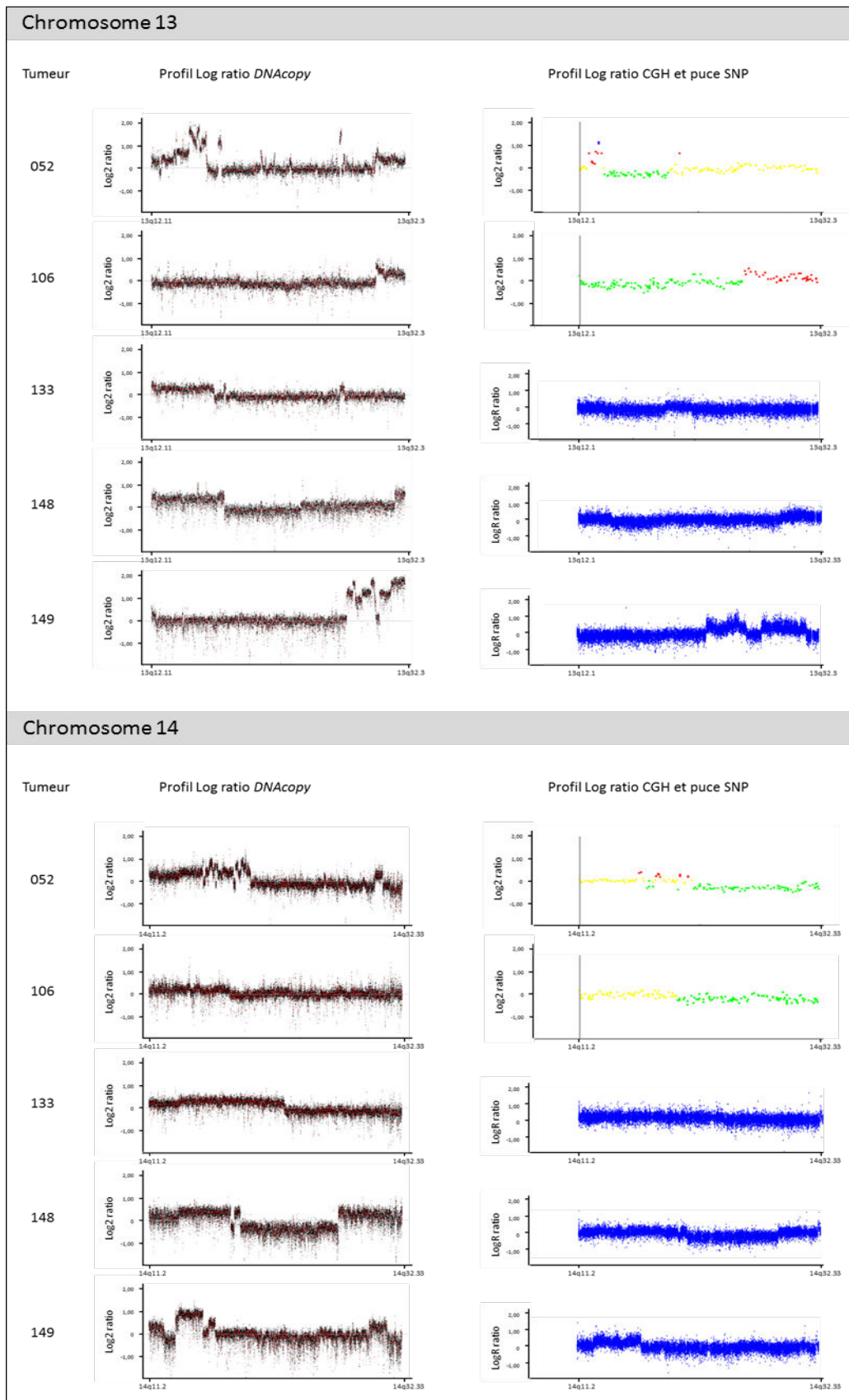
A l'aide des données de CNA, nous avons recherché les gènes qui seraient la cible d'une délétion homozygote. Pour cela, nous avons étudié le nombre de copies d'ADN des régions génomiques délétées. Le nombre de copies est calculé à partir du log2 ratio. Ainsi, lorsque la profondeur de lecture est identique entre la tumeur et le constitutionnel, le ratio est égal à 1, le log2ratio est égal à 0 et le nombre de copies est égal à 2. Cette normalisation se fonde sur le caractère diploïde de l'ADN constitutionnel.

Nous avons sélectionné dans les zones minimales de délétion des chromosomes 13 et 14, les gènes dont le nombre de copies d'ADN calculé est inférieur à 0.8 (Table 14). Un nombre de copies égal à 0.8 correspond à un log2 ratio égal à -1.3253 et donc à un ratio  $\frac{\text{profondeur tumeur}}{\text{profondeur constitutionnel}}$  égal à 0.4. La profondeur de lecture de l'ADN tumoral au niveau de ces régions génomiques serait le reflet de la contamination de l'échantillon tumoral avec les cellules du stroma, le gène serait délété de façon homozygote dans la tumeur.

Sur le chromosome 13, une partie d'un exon ou plusieurs exons de 7 gènes sont présents à moins de 0.8 copie contre 15 gènes sur le chromosome 14. Les gènes *DLEU2* (chromosome 13) et *PAPLN* (chromosome 14) retiennent particulièrement notre attention, car sont la cible de CNA dans chacun des deux tumeurs.

Pour le gène *DLEU2*, la région 50.699.413-50.699.473 (hg19) contenant 60 nucléotides exoniques et la région 50.571.253-50.571.373 (hg19) contenant 120 nucléotides d'un exon sont présentes à moins de 0.8 copie respectivement dans la tumeur 106 et la tumeur 133. Pour le gène *PAPLN*, la région 73.726.114-73.726.294 (hg19) couvrant 140 nucléotides d'un exon est présente à 0.69 copie dans la tumeur 149, et les régions 73.711.348-73.712.891 (hg19) couvrant 2 exons et 73.729.355-73.730.360 (hg19) à cheval sur deux exons sont évaluées à 0.8 copie ou moins dans la tumeur 148.

D'après nos résultats, aucun gène n'est perdu de manière récurrente dans les tumeurs BRCA2 ou inactivé par une mutation ponctuelle dans une tumeur et la délétion homozygote dans les 4 autres tumeurs BRCA2. Cependant, il faut noter que le gène *PAPLN* est également la cible d'une délétion en phase de 30 nucléotides dans la tumeur 148. Les résultats de CNA n'ont pas été confirmés par une autre technique.



**Figure 17 : Représentation graphique du log<sub>2</sub> ratio des bras longs des chromosomes 13 et 14 obtenu par l'analyse des CNA à l'aide du package *DNAcopy* des données NGS de 5 tumeurs *BRCA2*. Les profils génomiques en Log ratio obtenus en *BAC-CGH* (tumeur 052 et 106) et puces *SNP* (tumeur 133, 148 et 149) sont présentés pour comparaison.**

Tumeur	Chromosome	Début 5' (hg 19)	Fin 3' (hg 19)	Log2Ratio	Nombre de copies	Gène
148	13	42017030	42017510	-1,9775	0,51	OR7E37P
149	13	43148413	43148653	-1,8575	0,55	TNFSF11
149	13	45915176	45915236	-1,5289	0,69	TPT1
149	13	46039025	46039364	-1,4079	0,75	COG3
133	13	50070512	50070632	-2,0151	0,49	PHF11
133	13	50571253	50571373	-1,4084	0,75	DLEU2;TRIM13
106	13	50699413	50699473	-1,6175	0,65	DLEU2
148	14	73423090	73423150	-1,4279	0,74	DCAF4
149	14	73711348	73712891	-1,3233	0,8	PAPLN
148	14	73726114	73726294	-1,544	0,69	PAPLN
149	14	73729355	73730360	-1,3374	0,79	PAPLN
149	14	73957788	73957908	-1,3783	0,77	C14orf169
106	14	74226618	74226858	-1,3475	0,79	ELMSAN1
149	14	74353541	74353661	-1,3363	0,79	ZNF410
149	14	74706163	74706403	-1,4277	0,74	VSX2
148	14	75179588	75179708	-1,6252	0,65	KIAA0317
149	14	77605524	77605644	-1,7685	0,59	ZDHHC22
148	14	77735552	77735672	-1,713	0,61	NGB
148	14	77744796	77745189	-1,6099	0,66	POMT2
148	14	91671048	91671168	-2,0739	0,48	C14orf159
148	14	91700339	91701119	-1,3891	0,76	GPR68
52	14	93389476	93389536	-1,682	0,62	CHGA
149	14	95916234	95916354	-1,3223	0,8	SYNE3
106	14	99947716	99947836	-1,5883	0,67	CCNK

**Table 14:** Liste de gènes présents dans la tumeur à moins de 0.8 copie dans les zones minimales de délétion des chromosomes 13 et 14. Les colonnes Début 5' et Fin 3' indiquent les bornes des régions délétées. Le Log2 Ratio est calculé selon le Log2 du ratio de la profondeur de lecture entre les deux échantillons tumeur et constitutionnel. Le nombre de copies est un nombre de copies calculé à partir du Log2 Ratio.



### 3.4. Variants identifiés et gènes candidats

#### 3.4.1. Confirmation des variants SNVs et INDELS identifiés

Par séquençage Sanger sur les prélèvements constitutionnels et tumoraux correspondants, nous avons confirmé la présence somatique exclusive des 10 SNVs (*BMP4*, *CTAGE5*, *ELF1*, *KATNAL1*, *FAM179B*, *FANCM*, *NID2*, *RABBGTA*, *TRAF3* et *UNC79*) et de 5 INDELS sur 6 (*FERMT2*, *MYH6*, *PAPLN*, *PRMT5* et *PCNX*). Les électrophorogrammes sont présentés en annexe 1 et 2.

Sur les 10 SNVs et 5 INDELS, seulement 2 SNVs et 2 INDELS appartiennent à la région minimale de délétion des chromosomes 13 et 14, le gène *ELF1* (chromosome 13), le gène *UNC79* (chromosome 14), le gène *PAPLN* (chromosome 14) et le gène *PCNX* (chromosome 14).

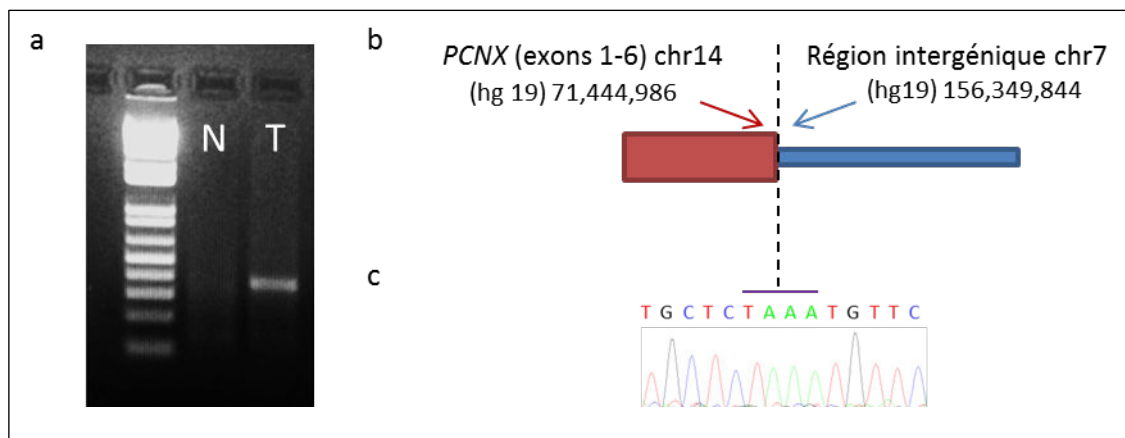
##### 3.4.1.1. Cas du variant du gène *DDHDI*

Nous n'avons pas confirmé par Sanger la nature somatique de la duplication du gène *DDHDI*. Ce variant a été retrouvé dans l'ADN tumoral mais aussi dans l'ADN constitutionnel. Nous pouvons expliquer par la présence de variants dans des séquences de mauvaise qualité que nous avons ignorées lors de la visualisation des fichiers .bam du constitutionnel. Ce variant n'est plus considéré comme candidat.

##### 3.4.1.2. Cas du variant du gène *PCNX*

L'INDEL du gène *PCNX* est décrite comme une insertion non bornée, seule une portion de 112 nucléotides de l'insertion est connue, et cette insertion est située dans l'exon 6 du gène. Nous avons essayé de séquencer l'insertion à l'aide d'amorces de PCR encadrant la position génomique du variant sans succès. Nous avons donc émis plusieurs hypothèses : l'INDEL n'est pas présente dans la tumeur, la séquence insérée est trop grande pour être séquencée en PCR classique ou symbolise en réalité un point de cassure lié à une translocation. Pour étayer l'hypothèse de la translocation, nous avons soumis la séquence insérée de 112 nucléotides à l'algorithme BLAT (cf. Table 13) [197] et avons trouvé une homologie de séquence de 100% sur le chromosome 7 entre les positions génomiques 156.349.733 et 156.349.844. Nous avons alors séquencé le point de cassure en dessinant une amorce sens (CATCGGACAGCTTCTGCC) localisée sur l'exon 6 du gène *PCNX* (chromosome 14) en amont du point de cassure et une amorce anti-sens (AAGCCACTGCACTCCAGC) localisée sur la région

intergénique du chromosome 7 entre les positions génomiques 156.349.733 et 156.349.844. La PCR réalisée avec ce couple d'amorce n'amplifie aucun fragment d'ADN dans l'échantillon constitutionnel, mais un produit de séquence est obtenu sur l'échantillon tumoral qui, après séquençage, confirme la présence de l'insertion somatique d'une portion du chromosome 7 au niveau des positions génomiques 156.349.844 et 71.444.986 des bras longs des chromosomes 7 et 14 (Figure 18). De plus on observe une homologie de séquence de 4 nucléotides TAAA entre la séquence insérée du chromosome 7 et le point de cassure sur le chromosome 14.



**Figure 18 : Etude de l'insertion d'une portion du bras long du chromosome 7 au niveau de l'exon 6 du gène *PCNX* localisé sur le chromosome 14 dans la tumeur BRCA2 133.** (a) La PCR sur l'ADN tumoral (T) et l'ADN constitutionnel (N) à l'aide d'une amorce Forward (CATCGGACAGCTTCTGCC) localisée sur le chromosome 14 et une amorce Reverse (AAGCCACTGCACTCCAGC) localisée sur le chromosome 7 confirme la présence de l'insertion somatique d'une portion du bras long du chromosome 7 au niveau de l'exon 6 du gène *PCNX*. (b) Schéma de la translocation entre l'exon 6 de *PCNX* et la région intergénique du chromosome 7. (c) Electrophérogramme de la séquence du produit de PCR confirmant le point de cassure de l'exon 6 du gène *PCNX*. La barre violette indique la séquence d'homologie TAAA du point de cassure.

Nous avons ensuite étudié le remaniement génomique en FISH sur la tumeur 133, en encadrant le gène *PCNX* sur le chromosome 14 et par conséquent le point de cassure localisé sur l'exon 6, à l'aide de sondes BlueFISH de la société Amplitech sur des empreintes de tumeur congelée. Deux BACs (RP11-137A13 et RP6-65G23, localisés respectivement en 71.067.776-71.258.426 et 71.127.903-71.294.089, pour augmenter le signal) marqués en SpectrumOrange (émet dans le rouge) en amont du gène *PCNX*, et 1 BAC (RP6-91H8 localisé en 71.576.994-71.705.191) en aval du gène marqué en SpectrumGreen (émet dans le vert) ont été utilisés (Figure 19A). D'après la construction, une translocation intragénique sur un allèle se symbolise par la présence d'un signal vert et d'un signal rouge isolé, un allèle normal se symbolise par la présence d'une fusion du signal vert et rouge. Le profil

majoritairement observé dans des noyaux intacts et non chevauchants est une fusion des signaux verts et rouge et un signal rouge isolé (Figure 19C).

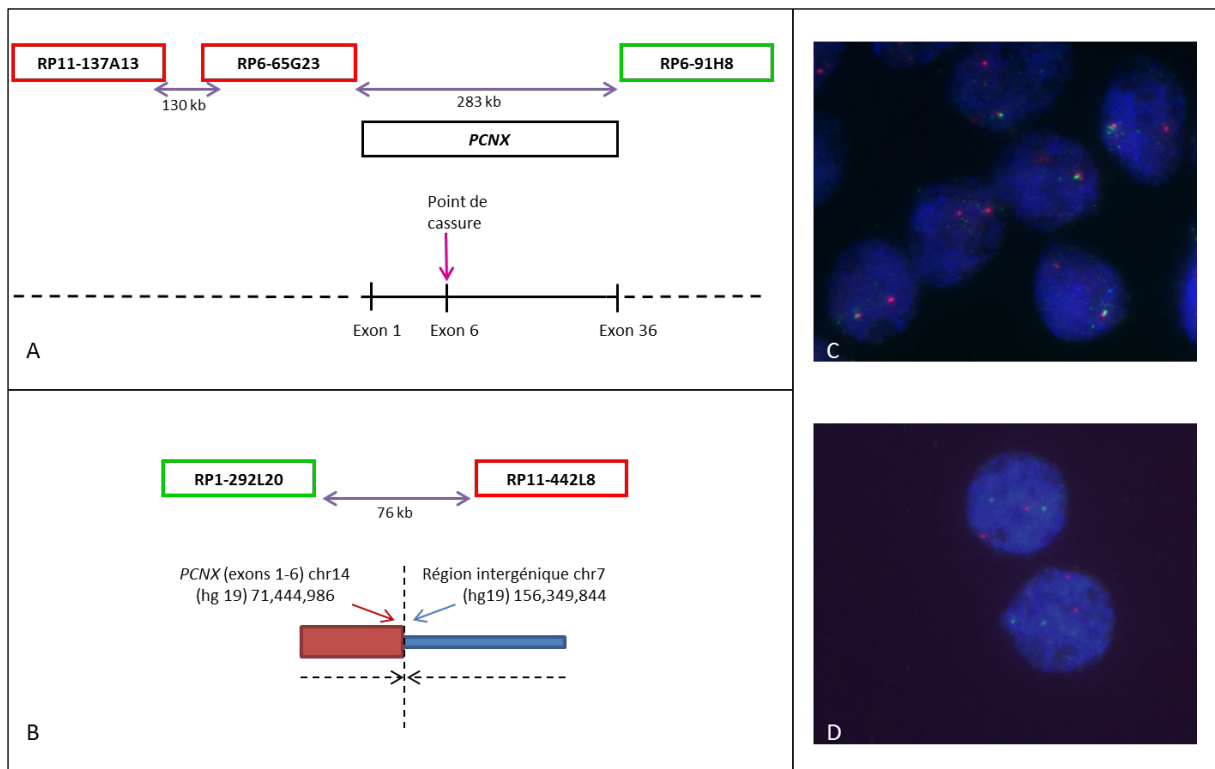
Ce résultat permet deux observations : (i) le signal rouge isolé et la présence d'un signal vert est en faveur d'une translocation déséquilibrée entraînant la présence de la région du chromosome 14 en 3' du point de cassure du gène *PCNX* (ii) la fusion du signal vert avec le signal rouge indique la présence d'un allèle normal. Effectivement, si le profil de la tumeur 133 est en faveur d'une délétion d'un allèle en amont du gène *PCNX* emportant le reste du bras long du chromosome 14, les résultats de NGS évoquent plutôt un point de cassure au niveau du gène *PCNX*. Notre hypothèse que les 2 allèles du gène *PCNX* étaient inactivés, l'un par la délétion et l'autre par la translocation est donc contredite par les résultats de FISH qui suggèrent que les deux événements concernent en fait le même allèle et que le point de cassure sur le chromosome 14 de la tumeur 133 en position 71.444.986 dans l'exon 6 du gène *PCNX* correspond à la borne centromérique de la délétion du chromosome 14. Ce point de cassure bornerait la région minimale de délétion car la tumeur 133 présente la plus petite région délétée du chromosome 14 (cf. Figure 7) et les données de la littérature localisent la borne centromérique en amont du gène *PCNX* (70.200.000 hg18) [190].

Nous avons réalisé une autre construction de FISH pour valider notre dernière observation et vérifier à la fois que le point de cassure de l'exon 6 de *PCNX* correspond à la borne 5' de la région délétée du bras long du chromosome 14 sur la tumeur 133 et qu'il y a eu translocation déséquilibrée entre les chromosomes 7 et 14. Pour cela, nous avons sélectionné un BAC (RP1-292L20 localisé en 71.315.515-71.439.535) complémentaire de la région en amont du point de cassure sur le chromosome 14 marqué en SpectrumOrange (émet dans le rouge) et 1 BAC (RP11-442L8 localisé en 156.162.651-156.327.120) en aval du point de cassure sur le chromosome 7 marqué SpectrumGreen (émet dans le vert) (Figure 19B). D'après la construction, la translocation intragénique entre les bras longs des chromosomes 7 et 14 entraînant la délétion du bras long du chromosome 14 se symbolise par la présence d'un signal rouge et d'un signal vert isolé (allèle normal) et par la fusion d'un signal vert et d'un signal rouge (allèle transloqué).

Cependant, le profil majoritairement observé dans des noyaux intacts et non chevauchants est 2 signaux verts isolés associés à 2 signaux rouges isolés (Figure 19D). L'interprétation des résultats confirme qu'en amont du point de cassure le chromosome 14 est diluée, ce qui signifie que le point de cassure observé sur le chromosome 14 de la tumeur 133 au niveau de l'exon 6 du gène *PCNX* correspond au début de la région délétée (borne 5'). La région minimale de délétion du chromosome 14 s'étend donc de 71.444.986 à 99.51.630.

En revanche, nous n'avons pas observé de fusion entre les BACs marqués en SpectrumOrange et SpectrumGreen, ce qui laisse penser que la translocation entre les chromosomes 7 et 14 ne concerne pas la totalité du chromosome 7. La région maximale du chromosome 7 impliquée dans la

translocation englobe la région située entre le point de cassure du chromosome 7 et la borne 3' de la BAC (car le brin du chromosome 7 transloqué est inversé) soit la région 156.327.120-156.349.844.



**Figure 19 : Etude du remaniement génomique au niveau du gène PCNX en FISH sur la tumeur 133.** (A) Deux BACs (RP11-137A13 et RP6-65G23) marqués en SpectrumOrange en amont du gène PCNX, et un BAC (RP6-91H8) en aval du gène marqué en SpectrumVert sont hybridés sur l'ADN dénaturé d'empreintes de tumeur congelée. (B) Un BAC (RP1-292L20) marqué en SpectrumVert en amont du gène PCNX, et un BAC (RP11-442L8) en aval du gène marqué en SpectrumOrange sont hybridés sur l'ADN dénaturé d'empreintes de tumeur congelée. Les distances entre les sondes BACs sont données en kilobases (kb). Les flèches en pointillé symbolisent l'orientation des chromosomes (centromère vers télomère). (C) Etude du point de cassure du gène PCNX sur la tumeur 133 à l'aide des BACs présentés dans la construction (A). (D) Etude du point de cassure du gène PCNX sur la tumeur 133 à l'aide des BACs présentées dans la construction (B). Dans l'expérience C, on observe la fusion entre un signal vert et un signal rouge ainsi qu'un signal rouge isolé. La fusion témoigne de la présence d'un allèle normal et le signal rouge isolé associé à la perte d'un signal vert symbolise la perte de la région 71.576.994-71.705.191 en aval du point de cassure étudié. Dans l'expérience D, on observe 2 signaux verts isolés associés à 2 signaux rouges isolés, ce qui signifie qu'en amont du point de cassure le chromosome 14 est diploïde et que la translocation entre les chromosomes 7 et 14 implique au maximum la région située entre le point de cassure du chromosome 7 et la borne 3' de la BAC (car le brin du chromosome 7 transloqué est inversé) soit la région 156.327.120-156.349.844.

### 3.4.2. Hiérarchisation des variants SNVs et INDELs identifiés

Après caractérisation des événements somatiques par séquençage Sanger et FISH, nous avons étudié les 10 variants SNVs et les 5 variants INDELs selon plusieurs critères comme le retentissement de la mutation sur la protéine, la localisation du gène candidat dans la zone minimale de délétion des chromosomes 13 et 14 et l'existence d'un lien entre la fonction de la protéine et le processus tumoral, dans le but de hiérarchiser les variants de signification inconnue (Table 15).

Tumeur	SNV		INDEL		CNA		
	Gène	Conséquence	Gène	Conséquence	Gène	Région perdue hg19	Nombre de copies
52	/		/		/		
106	/		<i>PRMT5</i>	c.1101_1144del;p.Gly369ProfsX24	<i>DLEU2</i>	50699413-50699473	0,65
			<i>FERMT2</i>	c.1436_1438dup;p.Thr479dup			
133			<i>PCNX</i>	cassure intragénique?	<i>DLEU2</i>	50571253-50571373	0,75
148	<i>KATNAL1</i>	c.1343C>T;p.Ser448Phe	<i>PAPLN</i>	c.3063_3092del;p.Pro1023_Gln1032del	<i>PAPLN</i>	73726114-73726294	0,69
	<i>RABGGTA</i>	c.1107A>T;p.Glu369Asp					
	<i>CTAGE5</i>	c.2198A>T;p.Asp733Val					
	<i>FAM179B</i>	c.1996C>A;p.Pro666Thr					
	<i>NID2</i>	c.1784C>A;p.Ala595Asp					
149	<i>ELF1</i>	c.1150C>T;p.Arg384Trp	<i>MYH6</i>	c.4612_4626del;p.Glu1538_Leu1542del	<i>PAPLN</i>	73711348-73712891	0,80
	<i>FANCM</i>	c.6079C>T;p.His202Tyr			<i>PAPLN</i>	73729355-73730360	0,79
	<i>BMP4</i>	c.718C>T;p.His240Tyr					
	<i>UNC79</i>	c.5623G>A;p.Val1875Met					
	<i>TRAF3</i>	c.1033G>C;p.Glu345Gln					

**Table 15: Tableau récapitulatif des différentes altérations identifiées dans les tumeurs BRCA2 étudiées en NGS.** Description des 10 SNVs somatiques et des 5 INDELs étudiés en Sanger ainsi que les CNA dont le nombre de copies est inférieur à 0.8. Les CNA présentés ciblent le même gène dans au moins deux tumeurs et sont localisés dans la région minimale de délétion des chromosomes 13 et 14.

Les 10 variants SNVs somatiques identifiés sont des variants faux-sens, aucun variant non-sens n'a été identifié. Les variations de type INDELs exoniques somatiques sont des insertions ou délétions en phase de triplés de nucléotides pour les gènes *MYH6*, *FERMT2* et *PAPLN*, ne les faisant pas considérer comme des mutations avec un retentissement majeur sur la protéine. En revanche, la délétion de 44 nucléotides dans l'exon 11 du gène *PRMT5* entraîne l'acquisition d'un codon stop prématuré ayant un retentissement probable majeur sur la fonction de la protéine. Cependant ce gène est en amont de la région minimale de délétion du chromosome 14, et d'après l'analyse des CNA, le second allèle du gène n'est pas délété dans la tumeur présentant ce variant, réfutant l'hypothèse d'une inactivation bi-allélique de ce gène.

Cinq variants faux-sens SNVs des gènes *FAM179B*, *FANCM*, *BMP4*, *TRAF3*, *ELF1* et deux variants INDELS des gènes *FERMT2* et *PRMT5* constituent des gènes candidats potentiels en raison de leur fonction probablement liée à la tumorigénèse. De plus, le gène *ELF1* est localisé dans la région minimale de délétion du chromosome 13.

Enfin, les gènes *PAPLN* et *UNC79* sont localisés dans la région minimale de délétion du chromosome 14, mais codent pour une protéine dont la fonction ne semble pas liée à la tumorigénèse.

Les variants faux-sens des gènes *KATNAL1*, *RABGGTA*, *CTAGE5*, *NID2* et *MYH6* dont la fonction de la protéine n'est pas liée au processus tumoral et qui ne sont pas localisés dans la région minimale de délétion ne seront pas à étudier dans un premier temps.

De plus, l'analyse des CNA montre une inactivation bi-allélique de deux régions distinctes du gène *DLEU2* respectivement dans deux tumeurs (106 et 133). Le gène *PAPLN* possède 3 régions exoniques évaluées à moins de 0.8 copie, une région dans la tumeur 148 et deux régions dans la tumeur 149. Il faut noter que la tumeur 148 est également la cible d'une délétion de 30 nucléotides dans une autre région génomique du gène *PAPLN*. Le gène *DLEU2*, comme le gène *PAPLN* sont localisés dans la région minimale de délétion du chromosome 13.

A l'issue de notre analyse, 8 gènes candidats dont le variant est d'origine somatique et codant pour des protéines dont la fonction serait liée à un processus tumoral sont sélectionnés. Ce sont les gènes *BMP4*, *DLEU2*, *ELF1*, *FAM179B*, *FANCM*, *FERMT2*, *PRMT5* et *TRAF3* (Table 16).

Tumeur	Chromosome	Position hg19	Variant	Nature	Conséquence	Gène	Description
149	13	41515163	SNV	faux-sens	c.1150C>T;p.Arg384Trp	<i>ELF1</i>	E74-like factor 1
148	14	45433620	SNV	faux-sens	c.1996C>A;p.Pro666Thr	<i>FAM179B</i>	family with sequence similarity 179, member B
149	14	45669143	SNV	faux-sens	c.6079C>T;p.His202Tyr	<i>FANCM</i>	Fanconi anemia, complementation group M
149	14	54417259	SNV	faux-sens	c.718C>T;p.His240Tyr	<i>BMP4</i>	bone morphogenetic protein 4
149	14	103369664	SNV	faux-sens	c.1033G>C;p.Glu345Gln	<i>TRAF3</i>	TNF receptor-associated factor 3
106	14	23393534	INDEL	délétion tronquante	c.1101_1144del;p.Gly369ProfsX24	<i>PRMT5</i>	protein arginine methyltransferase 5
106	14	53331283	INDEL	insertion en phase	c.1436_1438dup;p.Thr479dup	<i>FERMT2</i>	fermitin family member 2
106	13	50699413-50699473	CNA	nb de copies=0,65	délétion bi-allélique de 36 NT d'un exon	<i>DLEU2</i>	deleted in lymphocytic leukemia 2
133	13	50571253-50571373	CNA	nb de copies=0,75	délétion bi-allélique de 120 NT d'un exon	<i>DLEU2</i>	deleted in lymphocytic leukemia 2

**Table 16 : Description des variants somatiques SNVs, INDELS et CNA de signification inconnue des 8 gènes candidats finaux.**

### 3.4.3. Analyse bibliographique des gènes candidats en lien avec la tumorigénèse

#### 3.4.3.1. Fonction potentielle susceptive de tumeurs

##### *BMP4 (bone morphogenetic protein 4)*

Le variant du gène *BMP4* est situé dans l'exon 5 sur le domaine N-terminal c.718C>T; .His240Tyr. Ce gène est muté dans 0.23% des cancers du sein selon la base COSMIC [201]. La protéine BMP4 appartient à la famille du facteur de croissance TGF- $\beta$  (*transforming growth factor  $\beta$* ) et régule la prolifération cellulaire, la différenciation et la mobilité. Différentes études montrent que BMP4 est capable de réduire la croissance des cellules de lignées de cancers du sein tout en facilitant la migration et l'invasion. La surexpression de BMP4 dans la lignée MBA-MD-231 favorise la migration et l'invasion des cellules cancéreuses alors qu'une inhibition de l'expression de BMP4 par ARN interférence diminue la migration et l'invasion des cellules cancéreuses mais augmente la prolifération cellulaire [202]. Une étude de 2010 alloue à la protéine BMP4 un rôle régulateur dans la croissance cellulaire, la migration et l'invasion ainsi que des rôliétés à la fois susceptibles de tumeurs et oncogéniques comme TGF-  $\beta$  [203].

##### *ELF1 (E74-like factor 1)*

Le variant c.1150C>T; .Arg384Trp de l'exon 8 du gène *ELF1* est enregistré dans la base dbSNP (rs148475737). Sa fréquence dans la population générale selon les données du projet NGS NHLBI Exome Sequencing Project (ESP) est rare, et estimée à 0.0001 [198]. Ce gène est muté dans 0.31% des cancers du sein selon la base COSMIC [201]. Il appartient à la famille des facteurs de transcription ETS (pour *E26 transformation specific sequence*) qui contrôlent des gènes impliqués dans des processus cellulaires critiques, tels que la prolifération cellulaire, l'apoptose, l'hématopoïèse, la différenciation, l'angiogénèse et l'invasion [204]. Les protéines Ets possèdent toutes un domaine ETS capable de se lier à l'ADN (domaine hélice-tour-hélice) et reconnaissent une séquence GGAA/T. Certaines protéines Ets contiennent un domaine PNT important dans la reconnaissance protéine-protéine, ce domaine est absent de la protéine ELF1. Le rôle des protéines Ets dans la carcinogénèse est soutenu par le fait que les gènes *Ets* sont la cible de translocations et sont localisés sur des régions chromosomiques délétées ou que leur expression est altérée dans les tumeurs solides et leucémies [205;206]. Il a été montré que l'introduction du gène chimérique présent dans 85% des Sarcomes d'Ewing *EWS/FLI* (où EWS appartient à la famille des protéines ETS, code pour *Ewing Sarcoma Breakpoint Region 1*) ayant perdu la capacité de se lier à l'ADN en raison de mutations ou d'une

délétion du domaine Ets dans une lignée de fibroblastes murins, a pu accélérer la formation de tumeurs [207]. Cependant le variant identifié dans la tumeur BRCA2 de notre étude n'est pas localisé dans le domaine Ets. Cette famille module l'expression d'oncogènes et de gènes suppresseurs de tumeur par la régulation directe du promoteur ou par une interaction protéine-protéine [205]. La protéine ELF1 s'exprime préférentiellement dans les cellules hématopoïétiques et les érythrocytes en développement [208]. Elle régule la transcription des gènes inductibles durant l'activation des cellules T et se lie directement à la protéine Rb codée par le gène suppresseur de tumeur *Rb (Retinoblastoma)* pour former un complexe Elf-1-Rb non-hosphorylé dans les cellules T. Une surexpression de Rb non-phosphorylée entraîne l'inhibition de Elf-1 et son incapacité à activer l'expression des gènes dans les cellules T. Après l'activation des cellules T, la phosphorylation de Rb entraîne la dissociation des complexes Rb-Elf-1, permettant à Elf-1 de se lier et d'activer l'expression des gènes nécessaires à l'activation des cellules T [209]. Dans le cancer du sein, il a été montré que le gène *ELF1* augmente l'expression du promoteur ERBB2 [210]. L'étude de la région minimale du promoteur du gène *BRCA2* a identifié des séquences canoniques correspondantes aux motifs de reconnaissance du facteur de transcription *ELF1* et l'expression de *BRCA2* dépendante du cycle cellulaire a été associée à la liaison de Elf-1 sur le promoteur *BRCA2*. La liaison de Elf-1 sur le motif de reconnaissance Ets active le promoteur *BRCA2* lui conférant un rôle suppresseur de tumeur [211;212].

#### *DLEU2 (deleted in lymphocytic leukemia 2)*

Le cluster miR-15a~16-1 est codé dans un intron du long ARN R-codant appelé DLEU2. Dans les leucémies lymphoïdes chroniques, l'aberration génomique majeure est la délétion en 13q14.3 (55% des cas) [213], suggérant la présence d'un gène suppresseur de tumeur. Une région minimale de délétion inférieure à 300 kb qui comprend le gène *DLEU2* ainsi que le cluster miR-15a/16-1 localisé dans l'intron de *DLEU2* a été identifiée [214]. La délétion bi-allélique de la région minimale de délétion DLEU2/miR-15a~16-1 chez la souris augmente le développement de maladies clonales lymphoprolifératives dans le sang périphérique. La délétion exclusive du cluster miR-15a~16-1 accélère la prolifération des cellules B par la modulation des gènes contrôlant la progression du cycle cellulaire [215]. D'après ces différents résultats, *DLEU2* serait un gène suppresseur de tumeur.

#### *FANCM (Fanconi anemia, complementation group M)*

Le variant faux-sens dans l'exon 23 du gène *FANCM* c.6079C>T; p.His2027Tyr n'est décrit dans aucune base de données de SNP. Ce gène est muté dans 0.68% des cancers du sein selon la base COSMIC [201]. La voie de l'anémie de Fanconi (FA) est impliquée dans la réparation au niveau des



fourches de répllication de l'ADN, et il semble que le complexe FANCM provoque l'activation de la voie à rés sa hosphorylation. L'inactivation de *FANCM* chez la souris entraîne un défaut partiel de la voie FA et une susceptibilité aux agents alkylants [216;217]. L'activité ADN-translocase de FANCM suggère qu'elle est capable de se déplacer sur une double hélice d'ADN, et qu'elle pourrait constituer un lien entre la chromatine et le complexe FA-E3 ligase [218]. L'activité E3 ligase du complexe est stimulée lors de la présence d'un pontage inter-brin au niveau de la fourche. De plus, FANCM est liée sur le plan fonctionnel à la protéine BRCA2. En effet, l'ubiquitinylation de FANCD2 est dépendante du domaine hélicase de FANCM et permet le recrutement de BRCA2 ainsi que la formation du complexe FANCD2-Ub-BRCA2 sur la chromatine [219;220]. Enfin, en plus de son rôle dans la translocation du complexe FA au niveau des fourches de répllication permettant l'ubiquitinylation, ces divers travaux suggèrent un rôle de FANCM dans la réparation de l'ADN [220;221].

### *TRAF3 (TNF receptor-associated factor 3)*

Le variant du gène *TRAF3* c.1033G>C; p.Glu345Gln est localisé sur l'exon 11. Ce gène est muté dans 0.35% des cancers du sein selon la base COSMIC [201]. La liaison de TRAF3 à la protéine membranaire CD40 membre de la famille du TNF (*tumor necrosis factor*) inhibe l'activation de la voie NF-κB [222]. Or le facteur de transcription NF-κB joue un rôle critique dans le développement et la progression du cancer du sein [223].

#### 3.4.3.2. Fonction potentielle oncogénique

### *FAM179B (family with sequence similarity 179, member B)*

Le gène *FAM179B* est muté dans 0.31% des cancers du sein selon la base COSMIC [201]. Il a récemment été décrit que les polymorphismes situés sur les sites de liaison aux microARNs ou sur des partenaires de la machinerie des microARNs peuvent affecter la fonction et le niveau d'expression des microARNs. On nomme ce groupe de SNP les miRSNPs [224]. Les miRSNPs ont récemment été impliqués dans la progression du cancer [225;226]. La fréquence du miRSNP rs1053667 localisé dans l'exon 19 du gène *FAM179b* en aval du codon stop est significativement plus grande dans les tumeurs, suggérant la contribution de ce miRSNP à la tumorigénèse [227]. En revanche, le variant c.1996C>A; p.Pro666Thr détecté dans notre étude n'a pas été caractérisé précédemment et le rs1053667 n'est pas présent dans l'ADN tumoral des 5 tumeurs BRCA2 ou le constitutionnel analysé.

### *FERMT2 (fermitin family member 2)*

Le variant du gène *FERMT2* c.1436\_1438del; .Thr479del est localisé dans l'exon 12. Ce gène est muté dans 0.10% des cancers du sein selon la base COSMIC [201]. *FERMT2* est membre de la famille des Fermitines impliquée dans l'activation de l'intégrine et la régulation de l'adhésion cellulaire des cellules de la matrice cellulaire. De plus, un lien entre *FERMT2* et le cancer du sein a été décrit [228]. Cette protéine favorise la prolifération des cellules cancéreuses mammaires, augmente la survie cellulaire en régulant négativement l'apoptose et induit la croissance tumorale [229]. De plus la surexpression de *FERMT2* dans la lignée MCF-7 conduit à l'instabilité génomique des cellules.

### *PRMT5 (protein arginine methyltransferase 5)*

Le variant du gène *PRMT5* c.1101\_1144del; .Gly369ProfsX24 est localisé dans l'exon 11 codant pour le domaine méthyl transférase. Cette mutation crée un décalage du cadre de lecture et un codon stop prématuré. Ce gène est muté dans 0.41% des cancers du sein selon la base COSMIC [201]. La protéine *PRMT5* a été identifiée comme cofacteur altérant les fonctions de la protéine *PDCD4 (Programme cell death 4)* dans le cancer du sein. Le gène *PDCD4* a été décrit comme gène suppresseur de tumeur, sa surexpression dans le cancer du poumon et de l'ovaire par exemple est corrélée à un bon pronostic [230]. L'expression de *PDCD4* est diminuée dans les cancers du sein canaux [231]. Cependant, la surexpression de *PDCD4* dans certains cancers est corrélée à une mauvaise survie, suggérant que des facteurs oncogéniques peuvent moduler ou inhiber la fonction de *PDCD4*. En effet, la coexpression de *PDCD4* et *PRMT5* dans un modèle orthotopique murin de cancer du sein accélère la croissance cellulaire et cette croissance est dépendante de l'activité catalytique de *PRMT5* [230]. La protéine *PRMT5* semble être un facteur tumoral dans un contexte de surexpression.

# **CONCLUSION GENERALE, DISCUSSION ET PERSPECTIVES**

---

## 4. DISCUSSION ET PERSPECTIVES

---

### 4.1. Synthèse des résultats obtenus

Cette étude montre que la délétion des bras longs des chromosomes 13 et 14 constitue un remaniement génomique fréquemment associé aux tumeurs BRCA2. Nous avons développé une signature basée sur les données d'expression qui distingue les tumeurs BRCA2 des tumeurs BRCA1. Le *clustering* hiérarchique des 66 gènes de la signature, qui regroupe les échantillons en fonction de l'intensité d'expression des gènes, sépare les tumeurs en deux groupes à la fois sur le set d'entraînement et le set de validation. Il semble donc que l'expression de ces gènes permette de classer les tumeurs en deux groupes, les tumeurs liées à *BRCA2* et les tumeurs BRCA1.

Une approche supplémentaire utilisant les résultats de profil génomique retrouve de manière spécifique aux tumeurs BRCA2 une association de délétion de matériel génétique au niveau des bras longs des chromosomes 13 et 14. Les régions retrouvées délétées dans les 9 tumeurs BRCA2 de l'étude, localisées entre les régions 13q13.3-13q14.3 et 14q24.2-14q32.2, ont été confirmées par plusieurs études indépendantes [139;142-145;190] et pourraient renseigner sur le statut BRCA2.

La délétion isolée des bras longs des chromosomes 13 et 14 a été décrite dans certaines études précédemment citées chez des contrôles, mais l'association de délétion des bras longs des chromosomes 13 et 14 renforce le caractère spécifique de ce remaniement. Une récente étude présente également la co-occurrence de la délétion 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 contrôle de statut BRCA2 négatif ou inconnu [190]. Dans cette même étude, la sensibilité est de 100% dans le groupe de validation contenant 7 tumeurs BRCA2 et 60 tumeurs contrôles contre une spécificité de 88%. Sur la totalité des CNA observés dans les tumeurs BRCA2, la délétion d'hétérozygotie au locus 14q32 est la plus spécifique des tumeurs BRCA2. La base de données GISTIC (*Tumorscape Release 1.6*) [232] rapporte également que la délétion de régions chromosomiques localisées sur le bras long du chromosome 13 (hg18 chr13 :44680312-57088104, 57088104-114059427, 18097312-46301361 et 50901262.-114059427) est un événement récurrent

dans les cancers du sein. En revanche, aucun enrichissement en régions délétées localisées sur le bras long du chromosome 14 n'est représenté dans cette base de donnée ( $q\text{-value} < 0.25$ ).

L'analyse GSEA montre un enrichissement en gènes sous-exprimés localisés au niveau de segments génomiques délétés de manière récurrente dans les tumeurs BRCA2, probablement lié à la réduction du nombre de copies d'ADN des régions délétées. Ceci indiquerait un effet de dosage génique au moins partiel dans la constitution de la signature BRCA2. La détection d'une délétion par rapport à une amplification sur des données d'expression est plus difficile en raison de la faible diminution d'intensité de signal jusqu'à atteindre le bruit de fond lors d'une délétion alors qu'il peut être multiplié par 100 dans le cas d'une amplification. Cette difficulté pourrait expliquer pourquoi d'une part, seuls certains gènes localisés sur les régions délétées des chromosomes 13 et 14 sont présents dans la liste des gènes sous-exprimés de la signature, et d'autre part pourquoi les signatures des tumeurs BRCA2 sur des données d'expression proposées par des travaux antérieurs n'ont pas rapporté la délétion des chromosomes 13 et 14 comme événement discriminant [146;147]. De plus, il est possible que la protéine BRCA2 n'apparaisse pas dans la signature en raison de la sensibilité insuffisante des sondes Affymetrix (le signal mesuré pour BRCA2 est proche du bruit de fond et montre une très faible variation). Pour documenter l'effet de dosage génique, il serait intéressant d'étudier l'expression des gènes de la signature transcriptionnelle par une technique de PCR en temps réel comme la méthode TaqMan.

Ainsi, l'association des différentes techniques étudiant le transcriptome et le génome des tumeurs BRCA2 permet de définir une liste de gènes différentiellement exprimés dans les tumeurs BRCA2 et BRCA1, et d'identifier une région commune de délétion sur les bras longs des chromosomes 13 et 14, fortement représentés dans la signature transcriptionnelle.

Afin d'essayer d'identifier les gènes cibles de ces délétions spécifiques des tumeurs BRCA2, l'étude a été poursuivie par un séquençage haut débit de cinq tumeurs BRCA2 et de l'ADN constitutionnel associé. Nous avons identifié comme altérations acquises dans l'ADN tumoral un total de 10 SNVs faux-sens, 5 INDELs dont une délétion de 44 nucléotides aboutissant à la synthèse d'une protéine tronquée, et 5 délétions à l'état homozygote. Aucun gène ne présente une de ces altérations dans plus d'une tumeur à l'exception de *PAPLN* et *DLEU2* altérés l'un et l'autre dans 2 tumeurs différentes mais sous-forme de CNA dont la fiabilité est moindre que les mutations ponctuelles. Aucun variant somatique exonique n'est localisé sur les gènes sous-exprimés de la signature transcriptionnelle des tumeurs BRCA2.

Ces gènes candidats ont été subdivisés en deux groupes. Le premier groupe rassemble les gènes codant pour une protéine ayant une fonction potentielle dans la tumorigénèse. Il contient 5 variants SNVs faux-sens (gènes *BMP4*, *ELF1*, *FAM179B*, *FANCM*, *TRAF3*), deux variants INDELs (*FERMT2* et *PRMT5*) dont un entraînant l'acquisition d'un codon stop prématuré (gène *PRMT5*) et la délétion

homozygote d'une partie du gène *DLEU2* retrouvée dans deux tumeurs. Dans ce groupe, les gènes *ELF1* et *DLEU2* sont également localisés dans la région minimale de délétion.

Le second groupe rassemble les gènes dont la fonction de la protéine associée ne semble pas liée au processus tumoral. Nous y retrouvons les 5 variants faux-sens (gènes *KATNAL1*, *RABGGTA*, *CTAGE5*, *NID2* et *UNC79*), 2 variants INDELS en phase (gènes *MYH6* et *PAPLN*) et à nouveau le gène *PAPLN* qui semble être la cible d'une délétion homozygote pour une portion du gène.

Ainsi, l'étude en NGS des bras longs des chromosomes 13 et 14 de cinq tumeurs BRCA2 et de leur ADN constitutionnel à l'airé n'a pas permis de mettre en évidence un événement somatique récurrent commun aux tumeurs BRCA2. Aucun variant non-sens d'origine somatique n'a été confirmé mais la délétion somatique de 44 nucléotides dans l'exon 11 du gène *PRMT5* générant un codon stop prématuré a été caractérisée dans une tumeur.

Nous avons sélectionné les variants ayant la plus forte probabilité d'être liés au cancer en priorisant les gènes impliqués dans la survie cellulaire, la prolifération et la progression du cycle cellulaire. A l'issue de notre analyse, nous avons identifié 8 gènes candidats (gènes *BMP4*, *DLEU2*, *ELF1*, *FAM179B*, *FANCM*, *FERMT2*, *PRMT5*, *TRAF3*) qui restent à étudier dans une large série de cas BRCA2 et de cas contrôles pour conclure sur leur contribution à la prolifération tumorale des cancers du sein liés à BRCA2.

## 4.2. Discussion et Perspectives

### 4.2.1. Application en pratique clinique des résultats obtenus

La spécificité des pertes de matériel chromosomique concernant les bras 13q et 14q ainsi que leur grande fréquence dans les tumeurs BRCA2, puisque décrites par de nombreuses études, offre une perspective clinique intéressante pour le diagnostic des tumeurs liées à *BRCA2*. Actuellement, l'indication d'une recherche de mutation dans *BRCA2* repose essentiellement sur des critères généalogiques de récurrence familiale de cancers du sein et de l'ovaire ((i) au moins 3 cas de cancers du sein chez des femmes apparentées au premier ou au second degré, ou (ii) au moins un cas de cancer du sein et un cas de cancer de l'ovaire, ou 2 cas de cancers du sein dont au moins un avant 36 ans, chez des femmes apparentées au premier degré (ou deuxième degré via un homme)), mais la fréquence de mise en évidence d'une mutation selon ces critères reste faible [182]. Selon le rapport de l'Inca de 2011, 4.9% des analyses du gène *BRCA2* ont conduit à l'identification d'une mutation [183]. De plus l'analyse moléculaire est laborieuse en raison de la taille de la séquence et de l'absence de mutations récurrentes, cependant le développement du NGS dans les laboratoires d'oncogénétique tend à diminuer le temps de l'analyse. La recherche en routine des mutations de *BRCA2* sur biospécimens tumoraux par NGS n'est pas actuellement réalisable pour la plupart des laboratoires d'oncogénétique en raison du coût et de la difficulté de l'analyse inhérente au caractère tumoral de l'échantillon (hétérogénéité aléatoire, multiclonalité, cellularité parfois insuffisante).

En attendant la mise en place d'une telle organisation, nous proposons un test FISH applicable en pratique médicale courante susceptible d'identifier les tumeurs BRCA2 par délétion des segments localisés sur les chromosomes 13 et 14. Ce test FISH pourrait être réalisé concomitamment à la recherche d'amplification du gène *ERBB2* dans les tumeurs du sein par cette même technique.

Avant la mise en place d'un tel protocole, une étude de sensibilité et de spécificité de ce test sur une grande série de tumeurs dont le statut constitutionnel *BRCA2* est connu permettra d'en évaluer l'intérêt diagnostique.

L'avancée médicale dans la recherche de thérapie ciblée des tumeurs BRCA2 rend nécessaire l'identification rapide de ces tumeurs. En particulier, dans les cellules déficientes en BRCA2, le défaut de réparation par recombinaison homologue associée à l'inhibition de PARP1 (poly-ADP-ribose polymerase 1) induit de façon sélective et massive la mort cellulaire selon un mécanisme de synergie létale appelée aussi létalité synthétique de l'anglais *synthetic lethality* [176;178].

Lors de la validation de la signature transcritomique sur une cohorte de 19 tumeurs BRCA2 et 12 tumeurs BRCA1, 5 tumeurs BRCA1 partagent la même signature moléculaire que les tumeurs

BRCA2. La similarité du profil d'expression entre ces 5 tumeurs BRCAX et les tumeurs BRCA2 laisse envisager que ces tumeurs BRCAX possèdent les mêmes caractéristiques moléculaires que les tumeurs BRCA2 et que cette similarité pourrait être en lien ou non avec une mutation du gène *BRCA2* dont la recherche entreprise sur l'ADN constitutionnel aurait échoué. Tester la sensibilité aux inhibiteurs de PARP de ces 5 tumeurs BRCAX serait intéressant. En effet, des essais cliniques suggèrent que l'activité des inhibiteurs de PARP n'est pas restreinte aux tumeurs liées à des mutations *BRCA1/2*. Les inhibiteurs de PARP pourraient avoir une action dans les tumeurs déficientes en d'autres gènes de réparation de l'ADN tels que *RAD51* et *FANCM* [233].

L'utilisation de ce test FISH caractérisant la co-délétion des bras longs des chromosomes 13 et 14 permettrait à la fois de renseigner sur le statut BRCA2 dans un but diagnostique mais aussi dans un intérêt thérapeutique pour sélectionner les tumeurs qui correspondraient aux indications des antiPARP.

#### 4.2.2. Signification biologique des remaniements génomiques

Une explication de ces remaniements récurrents est que la perte de fonction de BRCA2 entraîne une réponse alternative aux dommages de l'ADN toxique pour la cellule en présence de gènes localisés sur les segments délétés. L'inactivation de ces gènes permettrait à la cellule de progresser dans le cycle cellulaire et d'entrer dans un processus de cancérogénèse. Ce modèle suppose que le facteur génétique code pour une protéine impliquée dans la réparation de l'ADN ou dans le contrôle du cycle cellulaire.

##### 4.2.2.1. Cible de la perte du bras long du chromosome 13

Dans notre étude, la fraction délétée du bras long du chromosome 13 s'étend entre les régions 13q13.3 et 13q14.3 et n'embrasse pas le locus du gène *BRCA2* situé en 13q13.1 pour une unique tumeur (tumeur 43). Aussi, le gène suppresseur de tumeur *RBI* (13q14) pourrait être une cible potentielle de la délétion du chromosome 13. Aucune altération de la séquence nucléotidique du gène *RBI* n'a été caractérisée sur les 5 tumeurs analysées, or ce gène est un gène suppresseur de tumeur, ce qui implique qu'il doit être inactivé de manière bi-allélique. *RBI* ne semble pas être la cible de la perte du bras long du chromosome 13.

L'étude de Pécuchet *et al.*, rapporte une perte d'hétérozygotie au locus *BRCA2* dans 27/29 tumeurs BRCA2, soit 93%. Il paraît alors impossible d'exclure une contribution de BRCA2 à ce remaniement génomique. Dans le cas de la tumeur 43, la délétion d'une fraction du chromosome 13, même distale par rapport au gène *BRCA2* peut être due à la proximité du gène *BRCA2*. Il est possible que cette région soit perdue lors d'un réarrangement génomique complexe lié ou médié par BRCA2.



#### 4.2.2.2. Cible de la délétion du bras long du chromosome 14

##### a. Haploinsuffisance et allèle hypomorphe

La signification biologique de ces remaniements récurrents pourrait être la nécessité de l'haploinsuffisance d'un gène ou d'un groupe de gènes qui favoriseraient la progression tumorale liée à BRCA2. Dans ce modèle, les gènes suspects de tumeur localisés sur les régions minimales de ces segments délétés partagées par les 9 tumeurs BRCA2 de notre étude sont candidats. Les clones de cellules tumorales dans lesquelles un ou plusieurs gènes suspects de tumeurs sont haploinsuffisants émergent et peuvent progresser dans la carcinogénèse. Ce phénomène a précédemment été décrit chez la souris par l'étude des doubles mutants BRCA1<sup>-/-</sup> qui conduisent à la mort au stade de l'embryon. Cependant, l'haploinsuffisance ou l'absence d'expression de Chk2 par l'introduction d'une mutation inactivatrice dans la séquence de *Checkpoint Kinase 2* permet aux embryons de souris d'échapper à la mort causée par la déficience en BRCA1 et augmente le développement de tumeurs du sein après la naissance [234]. Une autre explication serait la présence d'un allèle hypomorphe sur ces fragments délétés, diminuant ainsi l'expression de la protéine. Tout comme l'haploinsuffisance, la simple délétion d'un allèle suffirait à perturber la fonction de la protéine.

D'après la base de donnée COSMIC, les gènes impliqués dans le cancer localisés sur la région régulièrement perdue du chromosome 14 dans les tumeurs BRCA2 sont les gènes *RAD51L1* (14q23-q24.2), *GPHN* (14q24), *NIN* (14q24), *TSHR* (14q31), *TRIP11* (14q31-q32), *BCL11B* (14q32.1), *TCL1A* (14q32.1), *TCL6* (14q32.1) et le gène suspect de tumeur *DICER1* (14q32.13) (Table 17).

Gène	Description	Bande chromosomique
<i>RAD51L1</i>	RAD51-like 1 ( <i>S. cerevisiae</i> ) (RAD51B)	14q23-q24.2
<i>GPHN</i>	gephyrin (GPH)	14q24
<i>NIN</i>	ninein (GSK3B interacting protein)	14q24
<i>TSHR</i>	thyroid stimulating hormone receptor	14q31
<i>TRIP11</i>	thyroid hormone receptor interactor 11	14q31-q32
<i>BCL11B</i>	B-cell CLL/lymphoma 11B (CTIP2)	14q32.1
<i>TCL1A</i>	T-cell leukemia/lymphoma 1A	14q32.1
<i>TCL6</i>	T-cell leukemia/lymphoma 6	14q32.1
<i>DICER1</i>	dicer 1; ribonuclease type III	14q32.13

**Table 17 : Gènes impliqués dans le cancer localisés sur la région minimale de délétion du chromosome 14. D'après la base COSMIC [201].**

b. Mode d'inactivation d'un gène su`resseur de tumeurs selon l'hy`othèse de Knudson

La sélection de ces `ertes de matériel en cas de déficience en BRCA2 suggère la `résence au niveau de ces segments chromosomiques d'un facteur génétique ou d'un grou`e de gènes suscep`tibles de s'o`oser à la cancérisation mammaire liée à BRCA2.

Selon l'hy`othèse de Knudson, l'inactivation bi-allélique d'un gène su`resseur de tumeur localisé sur ces segments serait dû, d'une `art à la délétion d'un allèle (évènement récurrent observé en `uce SNP et CGH), et d'autre `art à une mutation `onctuelle du second allèle que nous avons recherchée en séquençage haut débit.

D'a`rès l'analyse bibliogra`hique sur la fonction des gènes candidats en lien avec la tumorigénèse localisés sur le chromosome 14, les gènes ayant une `otentielle fonction su`ressive de tumeur sont les gènes *BMP4*, *FANCM* et *TRAF3*. Ce`ependant, ces variants somatiques sont des mutations faux-sens ne `ermettant `as de `redire l'im`act sur la fonction de la `rotéine et devront faire l'objet d'une étude `our établir leur im`act sur la fonction `rotéique et la `rolifération tumorale.

#### 4.2.2.3. Présence de sites fragiles

Une autre ex`plication serait que les régions délétées contiennent des sites fragiles, régions chromosomiques connues `our être fréquemment affectées `ar des cassures suite à l'effondrement des fourches de ré`lication. La ré`aration de ces régions est ensuite assurée soit `ar recombinaison homologue (HR) soit `ar jonction de brins non-homologues (NHEJ). Il est connu que le mécanisme de recombinaison homologue (HR) joue un rôle majeur dans la résolution des cassures double-brin [235]. Des délétions et/ou des réarrangements chromosomiques dans `lusieurs ty`es de cellules tumorales ont été observés au niveau des sites fragiles [236]. Ce`ependant les sites fragiles décrits sur les chromosomes 13 et 14 à savoir FRA13A (13q13.2), FRA13B (13q21), FRA13C (13q21.2), FRA13D (13q32), FRA14A (14q21.2), FRA14B (14q23) et FRA14C (14q24.1) sont en dehors des régions minimales de délétion [237].

En l'absence de BRCA2, les cassures `résentes au niveau de ces sites fragiles seraient `lus difficilement ré`arables. En effet, BRCA2 est nécessaire au recrutement de RAD51 au niveau des sites de cassure `our initier la recombinaison homologue [163]. C'est l'hy`othèse que nous avons `rivilégiée `our la cassure intragénique du gène *PCNX* mais qui s'est révélée en fait corres`ondre au `oint de cassure de l'évènement délétionnel déjà connu concernant le `remier allèle. Une ré-analyse des données, à la recherche d'une variation du nombre de co`ies entre exons d'un même gène siégeant dans la région communément délétée, `ourrait a``orter des arguments en faveur de cette hy`othèse.

#### 4.2.3.Limites de l'étude

Nous devons rendre en considération certaines limites de notre étude. Le nombre d'individus sélectionnés pour le séquençage dans le but de trouver un évènement récurrent est faible. Malgré la détection de variants somatiques dans des gènes et tumeurs différents, il faut garder à l'esprit que la détection d'un variant somatique dans une tumeur signifie une fréquence allélique potentielle pour ce variant de l'ordre de 1 sur 10, soit 10%. Il a`rait alors nécessaire d'augmenter la cohorte des tumeurs BRCA2, et de comparer les résultats avec une cohorte de tumeurs BRCA1.

De plus, la résolution de gènes candidats résultant de la caractérisation de variants dans une collection donnée par séquençage haut débit est insuffisante. La documentation des gènes candidats du premier groupe dans une plus grande cohorte de cancers du sein au regard d'une collection contrôle ainsi qu'une étude de ségrégation sont à réaliser afin de statuer sur l'association de ces gènes aux tumeurs BRCA2. Afin de réaliser une étude exhaustive des variants qui ne nécessiterait que peu ou pas de sélection des variants, il semble nécessaire d'affiner la région minimale de délétion. Actuellement, les régions minimales de délétion du chromosome 13 et 14 mesurent respectivement 14Mb (Mégabases) et 29Mb et contiennent chacune 111 gènes et 208 gènes.

Pour cela, il faudrait augmenter la cohorte des tumeurs BRCA2 pour restreindre la région délétée, mais aussi utiliser les puces SNP pour optimiser les résultats des analyses de variation du nombre de copies. En effet les puces SNP permettent à la fois d'évaluer le nombre de copies mais aussi la perte d'hétérozygotie.

##### 4.2.3.1. Variants non sélectionnés à cause des filtres utilisés

L'échec dans la caractérisation de remaniements récurrents peut également s'expliquer par notre stratégie d'analyse et les limites de la technique.

Tout d'abord, nous avons filtré les variants sur leur localisation et leur conséquence. Nous avons étudié les variants somatiques en-dehors des régions 3' et 5' UTR (*untranslated regions*). Or ces régions sont considérées comme des séquences régulatrices et peuvent être la cible de mutations qui influencent le niveau d'expression de la protéine [238;239]. La région 3'-UTR joue un rôle dans la traduction, la localisation et la stabilité des ARNm. La région 5'-UTR possède de nombreux sites liant des protéines qui favorisent ou inhibent l'expression du gène concerné. Cette régulation peut s'exercer au niveau de la transcription, de la stabilité de l'ARNm, de la traduction ou de l'épissage [240].

Nous avons également supprimé les variants introniques qui étaient à plus de 4 nucléotides du site d'épissage et les variants iso-sémantiques s'ils n'étaient pas positionnés dans les sites donneurs (GT 5')/accepteur (AG 3') d'épissage. Or, des mutations introniques plus profondes que -4 et des variations iso-sémantiques à l'extérieur des sites accepteur AG et donneur GT ayant un retentissement sur

l'émission ont été décrites [155]. Pour être exhaustif, il faudrait étudier tous les variants iso-sémantiques, et élargir aux variants introniques peu profonds.

Nous avons également étudié les variants qui avaient une profondeur de lecture supérieure ou égale à 8, ce qui a éliminé 31 faux-sens et 2 variants non-sens (gène *SPERT* et *TTC9*). Par sécurité, nous avons tout de même séquencé en Sanger les 2 variants non-sens, et l'un d'eux, le variant du gène *SPERT* était présent dans la tumeur mais aussi dans le constitutionnel attendu (Annexe 3b). Or, les filtres utilisés pour l'analyse des variants l'avaient exclu pour l'échantillon constitutionnel car couvert moins de 8 fois et de mauvaise qualité. Le variant non-sens du gène *TTC9* n'a pas été retrouvé ni dans la tumeur, ni dans le constitutionnel attendu par séquençage Sanger.

Il est possible qu'un variant somatique récurrent dans les tumeurs BRCA2 soit dans les 31 faux-sens couverts moins de 8 fois éliminés dans notre analyse. Pour s'affranchir de cette éventualité, nous avons étudié les 31 variants faux-sens, aucun gène n'est altéré dans les 5 tumeurs.

#### 4.2.3.2. Variants non identifiés par l'analyse bioinformatique

Il est également envisageable que nous ayons perdu des informations durant la phase d'analyse bioinformatique à l'origine de faux négatifs. Il est possible que les *reads* contenant les variants d'intérêts aient été éliminés en raison d'une mauvaise qualité de lecture, ou alors en raison d'un alignement non optimal.

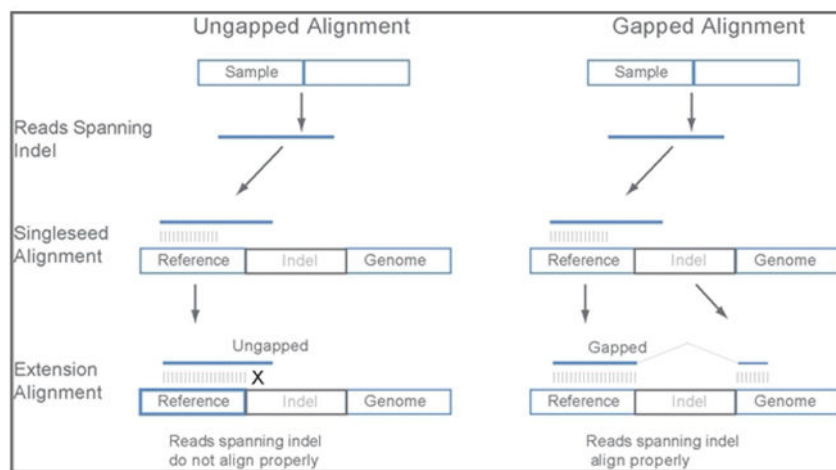
Nous avons alors repris les données .fastq et avons recommencé l'analyse bioinformatique dans le laboratoire en utilisant un pipeline que Casava1.8. Il s'agit du pipeline BWA-Picard-GATK-Annovar présenté en paragraphe 3.3.1.2 (cf. Figure 10).

Les 10 variants candidats de type SNV ont été retrouvés par cette méthode, en revanche aucun des 5 variants INDELS précédemment caractérisés n'a été identifié à l'aide de ce pipeline. Il semble que les paramètres utilisés lors de l'analyse bioinformatique sont à optimiser pour la détection des INDELS. Il est communément admis par les utilisateurs de NGS que la détection des INDELS est plus difficile que la détection des SNVs [241].

L'identification des INDELS est possible par l'utilisation d'un alignement autorisant les insertions et délétions, ce type d'alignement se nomme « *gapped alignment* ». Ces algorithmes d'alignement permettent d'aligner sur le génome la totalité d'un *read* découpé en segments. Un premier fragment est aligné, puis l'algorithme recherche pour le second fragment une autre région sur laquelle il s'aligne. La distance maximale séparant les deux régions est définie et constitue l'INDEL (Figure 20). L'algorithme d'alignement ELAND de Casava et l'algorithme BWA sont des *gapped alignments*, mais nous devons optimiser les paramètres de BWA pour augmenter la détection des INDELS.

Lors de l'analyse avec le nouveau pipeline, nous avons identifié 4 nouveaux variants non-sens localisés sur 3 gènes différents (*DGKH*, *KCTD12* et *MMP14*). Cependant, la visualisation des fichiers .bam de ces variants non-sens montre que la qualité des reads est médiocre et que l'on observe la même variation nucléotidique sur l'échantillon tumoral et l'échantillon constitutionnel concerné mais aussi dans les autres échantillons. Il est donc possible que ces variants soient d'origine constitutionnelle mais aussi que cela soit des artefacts (faux positif lié à un mauvais alignement), seul un séquençage Sanger pourrait statuer.

A l'aide du nouveau pipeline, nous avons également identifié 166 potentiels SNVs non-synonymes et 55 potentiels SNVs iso-sémantiques localisés sur 124 gènes différents dont 34 sur le chromosome 13 et 90 sur le chromosome 14. Ces variants seront à analyser selon la même procédure qui est décrite dans les résultats, à savoir la visualisation IGV des fichiers .bam, la sélection des variants lus 8 fois et le séquençage Sanger des variants qui ont passé ces filtres. Nous avons cherché dans ces variants, des variants localisés sur les gènes considérés comme candidats par l'analyse précédente (Table 11 et Table 13) : nous retrouvons 11 potentiels SNVs faux-sens localisés sur *ELF1*, *FANCM*, *PCNX* et *TRAF3*. Nous avons visualisé les variants de ces gènes sur IGV, mais nous observons le même phénomène que précédemment. Ces variants semblent être des artefacts.



**Figure 20 : Principe des algorithmes d'alignement de type « gapped »**, Extrait du *CASAVA v1.8.2 User Guide* [242]. Ces algorithmes d'alignement permettent d'aligner sur le génome la totalité d'un read découpé en segments. Un premier fragment est aligné, puis l'algorithme recherche pour le second fragment une autre région sur laquelle il s'aligne. La distance maximale séparant les deux régions est définie et constitue l'INDEL.

#### 4.2.3.3. Variants non identifiés par la technique utilisée

Pour favoriser la détection des variants structurelles (insertion, délétion et translocation) nous avons réalisé un séquençage paired-end non chevauchant qui identifie ces variations en comparant la distance observée entre la paire de *reads* et la distance théorique. Cependant, l'utilisation du RNA-seq (séquençage à ARN) permet d'apprécier le niveau d'expression des gènes à un moment donné dans un tissu, facilite la caractérisation de variants d'épissage et rend possible la détection de transcrits de fusion ou d'isoformes par rapport au séquençage de l'ADN génomique [243].

De plus le séquençage de l'ADN génomique n'a ciblé que les exons, ce qui signifie que nous n'avons pas séquencé les régions non codantes. Or le rôle des miRNA dans l'initiation et la progression tumorale a clairement été établi ces dernières années [65;66].

#### 4.2.4. Perspectives

La récurrence de ces remaniements génomiques dans les tumeurs du sein liées à BRCA2, dont la signification biologique n'a pas été clairement élucidée, impose de poursuivre l'investigation de manière plus approfondie. En effet, l'identification de mécanismes systématiques dans la cancérogénèse des tumeurs BRCA2 permet la recherche de cibles thérapeutiques potentielles, et offre ainsi la perspective d'une thérapie ciblée. La prise en compte des caractéristiques biologiques et génétiques de la tumeur, permet une prise en charge personnalisée et introduit la dimension moléculaire dans la stratégie thérapeutique.

# REFERENCES

---

## 4. REFERENCES

---

- (1) Boveri T. Zur Frage der Entstehung maligner Tumoren. Gustav Fisher Verlag, Jena; 1914.
- (2) Nowell PC. The minute chromosome (Ph1) in chronic granulocytic leukemia. *Blut* 1962 *A* r;8:65-6.
- (3) Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004 *Aug*;10(8):789-99.
- (4) Molist R, Remvikos Y, Dutrillaux B, Muleris M. Characterization of a new cytogenetic subtype of ductal breast carcinomas. *Oncogene* 2004 *Aug* 5;23(35):5986-93.
- (5) Rennstam K, Ahlstedt-Soini M, Baldetor B, et al. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res* 2003 *Dec* 15;63(24):8861-8.
- (6) Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992 *Oct* 30;258(5083):818-21.
- (7) Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* 2002 *Oct* 1;99(20):12963-8.
- (8) Seute A, Sinn HP, Schlenk RF, et al. Clinical relevance of genomic aberrations in homogeneously treated high-risk stage II/III breast cancer patients. *Int J Cancer* 2001 *Jul* 1;93(1):80-4.
- (9) Tirkkonen M, Tanner M, Karhu R, et al. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* 1998 *Mar*;21(3):177-84.
- (10) Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009 *Dec* 24;462(7276):1005-10.
- (11) Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptional architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012 *Jun* 21;486(7403):346-52.
- (12) Osborne C, Wilson P, Triathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. *Oncologist* 2004;9(4):361-77.
- (13) Kallioniemi A, Kallioniemi OP, Pileri J, et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* 1994 *Mar* 15;91(6):2156-60.
- (14) Roylance R, Gorman P, Harris W, et al. Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer. *Cancer Res* 1999 *A* r 1;59(7):1433-6.
- (15) Tavassoli FA, Devilee P. Pathology and Genetics. Tumours of the breast and female genital organs. World Health Organization Classification of Tumours 2003;60-76.



- (16) Adelaide J, Finetti P, Bekhouche I, et al. Integrated profiling of basal and luminal breast cancers. *Cancer Res* 2007 Dec 15;67(24):11565-75.
- (17) Kwek SS, Roy R, Zhou H, et al. Co-amplified genes at 8<sup>12</sup> and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. *Oncogene* 2009 Apr 30;28(17):1892-903.
- (18) Park K, Kwak K, Kim J, Lim S, Han S. c-myc amplification is associated with HER2 amplification and closely linked with cell proliferation in tissue microarray of nonselected breast cancers. *Hum Pathol* 2005 Jun;36(6):634-9.
- (19) Korkola J, Gray JW. Breast cancer genomes--form and function. *Curr Opin Genet Dev* 2010 Feb;20(1):4-14.
- (20) Rowley JD. The role of chromosome translocations in leukemogenesis. *Semin Hematol* 1999 Oct;36(4 Suppl 17):59-72.
- (21) Mitelman F, Johansson B, Mertens F (Eds.). *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* 2013 Available from: URL: <http://cgap.nci.nih.gov/Chromosomes/Mitelman>
- (22) Pandis N, Heim S, Bardi G, et al. Whole-arm t(1;16) and i(1q) as sole anomalies identify gain of 1q as a primary chromosomal abnormality in breast cancer. *Genes Chromosomes Cancer* 1992 Oct;5(3):235-8.
- (23) Pandis N, Jin Y, Gorunova L, et al. Chromosome analysis of 97 primary breast carcinomas: identification of eight karyotypic subgroups. *Genes Chromosomes Cancer* 1995 Mar;12(3):173-85.
- (24) Kokalj-Vokac N, Alemeida A, Gerbault-Seureau M, Malfroy B, Dutrillaux B. Two-color FISH characterization of i(1q) and der(1;16) in human breast cancer cells. *Genes Chromosomes Cancer* 1993 May;7(1):8-14.
- (25) Dutrillaux B, Gerbault-Seureau M, Zafrani B. Characterization of chromosomal anomalies in human breast cancer. A comparison of 30 trisomic cases with few chromosome changes. *Cancer Genet Cytogenet* 1990 Oct 15;49(2):203-17.
- (26) Dutrillaux B, Gerbault-Seureau M, Remvikos Y, Zafrani B, Prieur M. Breast cancer genetic evolution: I. Data from cytogenetics and DNA content. *Breast Cancer Res Treat* 1991 Nov;19(3):245-55.
- (27) Russnes HG, Vollan HK, Lingjaerde OC, et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med* 2010 Jun 30;2(38):38ra47.
- (28) Whitmore SA, Crawford J, Astolou S, et al. Construction of a high-resolution physical and transcription map of chromosome 16q24.3: a region of frequent loss of heterozygosity in sporadic breast cancer. *Genomics* 1998 May 15;50(1):1-8.
- (29) Kremmidiotis G, Baker E, Crawford J, et al. Localization of human cadherin genes to chromosome regions exhibiting cancer-related loss of heterozygosity. *Genomics* 1998 May 1;49(3):467-71.
- (30) Aulmann S, Blaker H, Penzel R, et al. CTCF gene mutations in invasive ductal breast cancer. *Breast Cancer Res Treat* 2003 Aug;80(3):347-52.
- (31) Berx G, Cleton-Jansen AM, Nollet F, et al. E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J* 1995 Dec 15;14(24):6107-15.

- (32) Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007 Apr;7(4):233-45.
- (33) Tognon C, Knezevich SR, Huntsman D, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* 2002 Nov;2(5):367-76.
- (34) Makretsov N, He M, Hayes M, et al. A fluorescence in situ hybridization study of ETV6-NTRK3 fusion gene in secretory breast carcinoma. *Genes Chromosomes Cancer* 2004 Jun;40(2):152-7.
- (35) Persson M, Andren Y, Mark J, et al. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc Natl Acad Sci U S A* 2009 Nov 3;106(44):18740-4.
- (36) Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 2012 Jun 21;486(7403):405-9.
- (37) Liu X, Baker E, Eyre HJ, Sutherland GR, Zhou M. Gamma-hergulin: a fusion gene of DOC-4 and neuregulin-1 derived from a chromosome translocation. *Oncogene* 1999 Nov 25;18(50):7110-4.
- (38) Schaefer G, Fitzpatrick VD, Sliwkowski MX. Gamma-hergulin: a novel heregulin isoform that is an autocrine growth factor for the human breast cancer cell line, MDA-MB-175. *Oncogene* 1997 Sep 18;15(12):1385-94.
- (39) Bohlander SK. ETV6: a versatile player in leukemogenesis. *Semin Cancer Biol* 2005 Jun;15(3):162-74.
- (40) Liu PP, Hajra A, Wijmenga C, Collins FS. Molecular pathogenesis of the chromosome 16 inversion in the M4Eo subtype of acute myeloid leukemia. *Blood* 1995 May 1;85(9):2289-302.
- (41) Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006 Oct 13;314(5797):268-74.
- (42) Samuels Y, Wang Z, Bardelli A, et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science* 2004 Apr 23;304(5670):554.
- (43) Carsten JD, Faber AL, Horn C, et al. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* 2007 Jul 26;448(7152):439-44.
- (44) Usary J, Llaca V, Karaca G, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004 Oct 7;23(46):7669-78.
- (45) Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 2010 Aug 12;466(7308):869-73.
- (46) Stephens PJ, Tarney PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012 Jun 21;486(7403):400-4.
- (47) Chuang JC, Jones PA. Epigenetics and microRNAs. *Pediatr Res* 2007 May;61(5 Pt 2):24R-9R.
- (48) Jenuwein T, Allis CD. Translating the histone code. *Science* 2001 Aug 10;293(5532):1074-80.

- (49) Qu G, Dubeau L, Narayan A, Yu MC, Ehrlich M. Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutat Res* 1999 Jan 25;423(1-2):91-101.
- (50) Dante R, Dante-Paire J, Rigal D, Roizes G. Methylation patterns of long interspersed repeated DNA and aliphatic repetitive DNA from human cell lines and tumors. *Anticancer Res* 1992 Mar;12(2):559-63.
- (51) Chalitchagorn K, Shuangshoti S, Hourai N, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* 2004 Nov 18;23(54):8841-6.
- (52) Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. *Biochem Biophys Res Commun* 1983 Feb 28;111(1):47-54.
- (53) Feinberg AP, Vogelstein B, Droller MJ, Baylin SB, Nelkin BD. Mutation affecting the 12th amino acid of the c-Ha-ras oncogene product occurs infrequently in human cancer. *Science* 1983 Jun 10;220(4602):1175-7.
- (54) Cheah MS, Wallace CD, Hoffman RM. Hypomethylation of DNA in human cancer cells: a site-specific change in the c-myc oncogene. *J Natl Cancer Inst* 1984 Nov;73(5):1057-65.
- (55) Paz MF, Fraga MF, Avila S, et al. A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res* 2003 Mar 1;63(5):1114-21.
- (56) Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002 Jun;3(6):415-28.
- (57) Robertson KD. DNA methylation and human disease. *Nat Rev Genet* 2005 Aug;6(8):597-610.
- (58) Mulero-Navarro S, Esteller M. Epigenetic biomarkers for human cancer: the time is now. *Crit Rev Oncol Hematol* 2008 Oct;68(1):1-11.
- (59) Tserga A, Michalopoulos NV, Levidou G, et al. Association of aberrant DNA methylation with clinicopathological features in breast cancer. *Oncol Rep* 2012 May;27(5):1630-8.
- (60) Brueckner B, Lyko F. DNA methyltransferase inhibitors: old and new drugs for an epigenetic cancer therapy. *Trends Pharmacol Sci* 2004 Nov;25(11):551-4.
- (61) Rocchi P, Tonelli R, Camerin C, et al. p21Waf1/Cip1 is a common target induced by short-chain fatty acid HDAC inhibitors (valproic acid, tributyrin and sodium butyrate) in neuroblastoma cells. *Oncol Rep* 2005 Jun;13(6):1139-44.
- (62) Kleer CG, Cao Q, Varambally S, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003 Sep 30;100(20):11606-11.
- (63) Stitt TN, Drujan D, Clarke BA, et al. The IGF-1/PI3K/Akt pathway prevents expression of muscle atrophy-induced ubiquitin ligases by inhibiting FOXO transcription factors. *Mol Cell* 2004 May 7;14(3):395-403.
- (64) Tan J, Yang X, Zhuang L, et al. Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev* 2007 May 1;21(9):1050-63.

- (65) Ma L, Teruya-Feldstein J, Weinberg RA. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 2007 Oct 11;449(7163):682-8.
- (66) Singh R, Mo YY. Role of microRNAs in breast cancer. *Cancer Biol Ther* 2013 Mar;14(3):201-12.
- (67) Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000 Aug 17;406(6797):747-52.
- (68) 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(19):10869-74.
- (69) Sorlie T. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur J Cancer* 2004 Dec;40(18):2667-75.
- (70) 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(14):8418-23.
- (71) Comprehensive molecular portraits of human breast tumours. *Nature* 2012 Oct 4;490(7418):61-70.
- (72) Solorzano CC, Middleton LP, Hunt KK, et al. Treatment and outcome of patients with intracystic papillary carcinoma of the breast. *Am J Surg* 2002 Oct;184(4):364-8.
- (73) Weigelt B, Geyer FC, Reis-Filho JS. Histological types of breast cancer: how special are they? *Mol Oncol* 2010 Jun;4(3):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(2):661-71.
- (75) Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 2006;7:96.
- (76) 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.
- (77) 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(2):429-33.
- (78) 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(10):4090-103.
- (79) Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a validation-based study. *Proc Natl Acad Sci U S A* 2003 Sep 2;100(18):10393-8.
- (80) Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular prognostic breast tumours by microarray analysis. *Oncogene* 2005 Jul 7;24(29):4660-71.
- (81) 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(5):R68.
- (82) 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(6871):530-6.

- (83) Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004 Dec 30;351(27):2817-26.
- (84) Dawson SJ, Rueda OM, Aricio S, Caldas C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J* 2013 Mar 6;32(5):617-28.
- (85) Broca P. *Traité des tumeurs*. P. Asselin 1866
- (86) Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971 Apr;68(4):820-3.
- (87) Knudson AG. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 2001 Nov;1(2):157-62.
- (88) Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet* 1955 Sep;7(3):277-318.
- (89) 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(5182):66-71.
- (90) Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990 Dec 21;250(4988):1684-9.
- (91) Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995 Dec 21;378(6559):789-92.
- (92) Rieger T, Gadzicki D, Meindl A, Schlegelberger B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* 2009 Jun;17(6):722-31.
- (93) Thomson D, Easton D. The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia* 2004 Jul;9(3):221-36.
- (94) Lalloo F, Evans DG. Familial breast cancer. *Clin Genet* 2012 Aug;82(2):105-14.
- (95) Ratajska M, Antoszewska E, Piskorz A, et al. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res Treat* 2012 Jan;131(1):89-97.
- (96) Erkkö H, Dowty JG, Nikkila J, et al. Penetrance analysis of the PALB2 c.1592delT founder mutation. *Clin Cancer Res* 2008 Jul 15;14(14):4667-71.
- (97) Thomson 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(1):95-9.
- (98) Pelttari LM, Heikkinen T, Thomson D, et al. RAD51C is a susceptibility gene for ovarian cancer. *Hum Mol Genet* 2011 Aug 15;20(16):3278-88.
- (99) Lu W, Wang X, Lin H, Lindor NM, Couch FJ. Mutation screening of RAD51C in high-risk breast and ovarian cancer families. *Fam Cancer* 2012 Sep;11(3):381-5.
- (100) Ghoussaini M, Fletcher O, Michailidou K, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* 2012 Mar;44(3):312-8.
- (101) Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007 Jun 28;447(7148):1087-93.

- (102) 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(7):870-4.
- (103) 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(7):865-9.
- (104) Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 11.2 and 14q24.1 (RAD51L1). *Nat Genet* 2009 May;41(5):579-84.
- (105) Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 2008 Jan;40(1):17-22.
- (106) Koonin EV, Altschul SF, Bork P. BRCA1 protein products ... Functional motifs. *Nat Genet* 1996 Jul;13(3):266-8.
- (107) 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(1):68-76.
- (108) Welsh PL, Owens KN, King MC. Insights into the functions of BRCA1 and BRCA2. *Trends Genet* 2000 Feb;16(2):69-74.
- (109) 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(20):11364-9.
- (110) Brzovic PS, Meza J, King MC, Klevit RE. The cancer-redising mutation C61G disrupts homodimer formation in the NH2-terminal BRCA1 RING finger domain. *J Biol Chem* 1998 Apr 3;273(14):7795-9.
- (111) 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(51):32863-8.
- (112) Huen MS, Sy SM, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. *Nat Rev Mol Cell Biol* 2010 Feb;11(2):138-48.
- (113) Prevost C, Takahashi M. Geometry of the DNA strands within the RecA nucleofilament: role in homologous recombination. *Q Rev Biophys* 2003 Nov;36(4):429-53.
- (114) 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(51):31941-4.
- (115) 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(5):1117-30.
- (116) Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. *Lancet* 2006 Feb 18;367(9510):595-604.
- (117) Thomson D, Easton DF. Cancer Incidence in BRCA1 mutation carriers. *J Natl Cancer Inst* 2002 Sep 18;94(18):1358-65.

- (118) van As`eren CJ, Brohet RM, Meijers-Heijboer EJ, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *J Med Genet* 2005 Se` ;42(9):711-9.
- (119) Cancer risks in BRCA2 mutation carriers. The Breast Cancer Linkage Consortium. *J Natl Cancer Inst* 1999 Aug 4;91(15):1310-6.
- (120) Thom`son D, Easton D. Variation in BRCA1 cancer risks by mutation `osition. *Cancer E`idemiol Biomarkers Prev* 2002 A`r;11(4):329-36.
- (121) Risch HA, McLaughlin JR, Cole DE, et al. Prevalence and `enetrance of germline BRCA1 and BRCA2 mutations in a `o`ulation series of 649 women with ovarian cancer. *Am J Hum Genet* 2001 Mar;68(3):700-10.
- (122) Collins N, McManus R, Wooster R, et al. Consistent loss of the wild ty`e allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12-13. *Oncogene* 1995 A`r 20;10(8):1673-5.
- (123) Boulton SJ. Cellular functions of the BRCA tumour-su``ressor `roteins. *Biochem Soc Trans* 2006 Nov;34(Pt 5):633-45.
- (124) Kinzler KW, Vogelstein B. Cancer-susce`tibility genes. Gatekee`ers and caretakers. *Nature* 1997 A`r 24;386(6627):761, 763.
- (125) Futreal PA, Liu Q, Shattuck-Eidens D, et al. BRCA1 mutations in `rimary breast and ovarian carcinomas. *Science* 1994 Oct 7;266(5182):120-2.
- (126) Cleton-Jansen AM, Collins N, Lakhani SR, et al. Loss of heterozygosity in s`oradic breast tumours at the BRCA2 locus on chromosome 13q12-q13. *Br J Cancer* 1995 Nov;72(5):1241-4.
- (127) Wilson CA, Ramos L, Villasenor MR, et al. Localization of human BRCA1 and its loss in high-grade, non-inherited breast carcinomas. *Nat Genet* 1999 Feb;21(2):236-40.
- (128) Russell PA, Pharoah PD, De FK, et al. Frequent loss of BRCA1 mRNA and `rotein ex`ression in s`oradic ovarian cancers. *Int J Cancer* 2000 Aug 1;87(3):317-21.
- (129) Welsh PL, King MC. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* 2001 A`r;10(7):705-13.
- (130) Baldwin RL, Nemeth E, Tran H, et al. BRCA1 `romoter region hy`ermethylation in ovarian carcinoma: a `o`ulation-based study. *Cancer Res* 2000 Oct 1;60(19):5329-33.
- (131) Collins N, Wooster R, Stratton MR. Absence of methylation of C`G dinucleotides within the `romoter of the breast cancer susce`tibility gene BRCA2 in normal tissues and in breast and ovarian cancers. *Br J Cancer* 1997;76(9):1150-6.
- (132) Hilton JL, Geisler JP, Rathe JA, et al. Inactivation of BRCA1 and BRCA2 in ovarian cancer. *J Natl Cancer Inst* 2002 Se` 18;94(18):1396-406.
- (133) Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and s`oradic cases. Breast Cancer Linkage Consortium. *Lancet* 1997 May 24;349(9064):1505-10.
- (134) Lakhani SR, Jacquemier J, Sloane JP, et al. Multifactorial analysis of differences between s`oradic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst* 1998 Aug 5;90(15):1138-45.

- (135) 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(1):121-8.
- (136) Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. *Nat Rev Cancer* 2004 Oct;4(10):814-9.
- (137) Chen J, Silver DP, Walitka 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(3):317-28.
- (138) 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(9):2310-8.
- (139) 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(4):R47.
- (140) Colombo M, Giarola M, Mariani L, et al. Cyclin D1 expression analysis in familial breast cancers may discriminate BRCA1 from BRCA2-linked cases. *Mod Pathol* 2008 Oct;21(10):1262-70.
- (141) 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(10 Pt 1):3606-14.
- (142) 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(17):7612-21.
- (143) 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(3):R42.
- (144) Jooisse 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(2):379-89.
- (145) 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(7):1222-7.
- (146) 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(3):661-77.
- (147) Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001 Feb 22;344(8):539-48.
- (148) Chen PL, Chen CF, Chen Y, et al. The BRCA2 repeats in BRCA2 are critical for RAD51 binding and resistance to methyl methanesulfonate treatment. *Proc Natl Acad Sci U S A* 1998 Apr 28;95(9):5287-92.
- (149) Yang H, Jeffrey PD, Miller J, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* 2002 Sep 13;297(5588):1837-48.



- (150) Mizuta R, LaSalle JM, Cheng HL, et al. RAB22 and RAB163/mouse BRCA2: proteins that specifically interact with the RAD51 protein. *Proc Natl Acad Sci U S A* 1997 Jun 24;94(13):6927-32.
- (151) Sharan SK, Bradley A. Identification and characterization of a microsatellite marker within murine Brca2 gene. *Mamm Genome* 1997 Jan;8(1):79.
- (152) Esashi F, Christ N, Gannon J, et al. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* 2005 Mar 31;434(7033):598-604.
- (153) Zou JP, Hirose Y, Siddique H, Rao VN, Reddy ES. Structure and expression of variant BRCA2a lacking the transactivation domain. *Oncol Rep* 1999 Mar;6(2):437-40.
- (154) 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(2):323-8.
- (155) Houdayer C, Caux-Moncoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012 Aug;33(8):1228-38.
- (156) Bieche I, Lidereau R. Increased level of exon 12 alternatively spliced BRCA2 transcripts in tumor breast tissue compared with normal tissue. *Cancer Res* 1999 Jun 1;59(11):2546-50.
- (157) Scully R, Livingston DM. In search of the tumour-suppressor functions of BRCA1 and BRCA2. *Nature* 2000 Nov 23;408(6811):429-32.
- (158) Sharan SK, Morimatsu M, Albrecht U, et al. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* 1997 Apr 24;386(6627):804-10.
- (159) Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 2002 Jan 25;108(2):171-82.
- (160) Vaughn JP, Cirisano FD, Huether G, et al. Cell cycle control of BRCA2. *Cancer Res* 1996 Oct 15;56(20):4590-4.
- (161) Marston NJ, Richards WJ, Hughes D, et al. Interaction between the product of the breast cancer susceptibility gene BRCA2 and DSS1, a protein functionally conserved from yeast to mammals. *Mol Cell Biol* 1999 Jul;19(7):4633-42.
- (162) Gudmundsdottir K, Lord CJ, Witt E, Tutt AN, Ashworth A. DSS1 is required for RAD51 focus formation and genomic stability in mammalian cells. *EMBO Rep* 2004 Oct;5(10):989-93.
- (163) Yang H, Li Q, Fan J, Holloman WK, Pavletich NP. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* 2005 Feb 10;433(7026):653-7.
- (164) Hussain S, Wilson JB, Medhurst AL, et al. Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum Mol Genet* 2004 Jun 15;13(12):1241-8.
- (165) Yamamoto K, Ishiai M, Matsushita N, et al. Fanconi anemia FANCG protein in mitigating radiation- and enzyme-induced DNA double-strand breaks by homologous recombination in vertebrate cells. *Mol Cell Biol* 2003 Aug;23(15):5421-30.

- (166) Nakanishi K, Yang YG, Pierce AJ, et al. Human Fanconi anemia monoubiquitination pathway promotes homologous DNA recombination. *Proc Natl Acad Sci U S A* 2005 Jan 25;102(4):1110-5.
- (167) Hughes-Davies L, Huntsman D, Ruas M, et al. EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 2003 Nov 26;115(5):523-35.
- (168) 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(6):719-29.
- (169) Milner J, Ponder B, Hughes-Davies L, Seltsmann M, Kouzarides T. Transcriptional activation functions in BRCA2. *Nature* 1997 Apr 24;386(6627):772-3.
- (170) Fuks F, Milner J, Kouzarides T. BRCA2 associates with acetyltransferase activity when bound to P/CAF. *Oncogene* 1998 Nov 12;17(19):2531-4.
- (171) Daniels MJ, Wang Y, Lee M, Venkitaraman AR. Abnormal cytokinesis in cells deficient in the breast cancer susceptibility protein BRCA2. *Science* 2004 Oct 29;306(5697):876-9.
- (172) Jonsdottir AB, Stefansson OA, Bjornsson J, et al. Tetraploidy in BRCA2 breast tumours. *Eur J Cancer* 2012 Feb;48(3):305-10.
- (173) Tutt A, Bertwistle D, Valentine J, et al. Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J* 2001 Sep 3;20(17):4704-16.
- (174) Tian XX, Rai D, Li J, et al. BRCA2 suppresses cell proliferation via stabilizing MAGE-D1. *Cancer Res* 2005 Jun 1;65(11):4747-53.
- (175) Barker PA, Salehi A. The MAGE proteins: emerging roles in cell cycle progression, apoptosis, and neurogenetic disease. *J Neurosci Res* 2002 Mar 15;67(6):705-12.
- (176) 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(7035):917-21.
- (177) Yu VP, Koehler M, Steinlein C, et al. Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Genes Dev* 2000 Jun 1;14(11):1400-6.
- (178) 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(7035):913-7.
- (179) 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(2):123-34.
- (180) Audeh MW, Carmichael J, Penson RT, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* 2010 Jul 24;376(9737):245-51.
- (181) Edwards SL, Brough R, Lord CJ, et al. Resistance to therapy caused by intragenic deletion in BRCA2. *Nature* 2008 Feb 28;451(7182):1111-5.
- (182) Moller P, Hagen AI, Auld J, et al. Genetic epidemiology of BRCA mutations--family history detects less than 50% of the mutation carriers. *Eur J Cancer* 2007 Jul;43(11):1713-7.
- (183) Synthèse de l'activité d'oncogénétique 2011. INCa . 2012.  
Ref Type: Magazine Article

- (184) Casilli F, Di Rocco ZC, Gad S, et al. Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments. *Hum Mutat* 2002 Sep;20(3):218-26.
- (185) Weber J, Miserere S, Champ J, et al. High-throughput simultaneous detection of point mutations and large-scale rearrangements by CE. *Electrophoresis* 2007 Dec;28(23):4282-8.
- (186) Houdayer C, Moncoutier V, Champ J, et al. Enhanced mismatch mutation analysis: simultaneous detection of point mutations and large scale rearrangements by capillary electrophoresis, application to BRCA1 and BRCA2. *Methods Mol Biol* 2010;653:147-80.
- (187) Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005 Oct 25;102(43):15545-50.
- (188) van Beers EH, van Welsem T, Wessels LF et al. Comparative Genomic Hybridization Profiles in Human BRCA1 and BRCA2 Breast Tumors Highlight Differential Sets of Genomic Aberrations. *Cancer Res* 2005;65:822–827.
- (189) Tischkowitz M, Xia B, Sabbaghian N, et al. Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci U S A* 2007 Apr 17;104(16):6788-93.
- (190) 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 Jun 10.
- (191) Andrews S. FastQC High Troughput Sequence QC Report. Babraham Bioinformatics 2012 May 3 Available from: URL: [www.bioinformatics.bbsrc.ac.uk/projects/fastqc](http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc)
- (192) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009 Jul 15;25(14):1754-60.
- (193) 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(9):1297-303.
- (194) Qin J. Duke University 2013 Available from: URL: <http://sites.duke.edu/sequencingatduke/illumina-quality-score-encoding/>
- (195) Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15;25(16):2078-9.
- (196) Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics* 2011 Aug 1;27(15):2156-8.
- (197) UCSC, BLAT Search Genome. *Genome Bioinformatics* 2013 Available from: URL: <http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>
- (198) Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle. WA 2013 September 10 Available from: URL: <http://evs.gs.washington.edu/EVS/>
- (199) Venkatraman ES, Olshen AB, Olshen E. DNACopy: DNA copy number data analysis. R package 1.26.0.

Ref Type: Generic

- (200) Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004 Oct;5(4):557-72.
- (201) COSMIC. Catalogue Of Somatic Mutations In Cancer 2013 September 1 Available from: URL: <http://www.sanger.ac.uk/cosmic>
- (202) Guo D, Huang J, Gong J. Bone morphogenetic protein 4 (BMP4) is required for migration and invasion of breast cancer. *Mol Cell Biochem* 2012 Apr;363(1-2):179-90.
- (203) Ketolainen JM, Alarmo EL, Tuominen VJ, Kallioniemi A. Parallel inhibition of cell growth and induction of cell migration and invasion in breast cancer cells by bone morphogenetic protein 4. *Breast Cancer Res Treat* 2010 Nov;124(2):377-86.
- (204) Hsu T, Trojanowska M, Watson DK. Ets proteins in biological control and cancer. *J Cell Biochem* 2004 Apr 1;91(5):896-903.
- (205) Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. *Gene* 2003 Jan 16;303:11-34.
- (206) Hsu T, Trojanowska M, Watson DK. Ets proteins in biological control and cancer. *J Cell Biochem* 2004 Apr 1;91(5):896-903.
- (207) Jaishankar S, Zhang J, Roussel MF, Baker SJ. Transforming activity of EWS/FLI is not strictly dependent upon DNA-binding activity. *Oncogene* 1999 Sep 30;18(40):5592-7.
- (208) Oettgen P, Alani RM, Barcinski MA, et al. Isolation and characterization of a novel epithelium-specific transcription factor, ESE-1, a member of the ets family. *Mol Cell Biol* 1997 Aug;17(8):4419-33.
- (209) Wang CY, Petryniak B, Thompson CB, Kaelin WG, Leiden JM. Regulation of the Ets-related transcription factor Elf-1 by binding to the retinoblastoma protein. *Science* 1993 May 28;260(5112):1330-5.
- (210) Scott GK, Chang CH, Erny KM, et al. Ets regulation of the erbB2 promoter. *Oncogene* 2000 Dec 18;19(55):6490-502.
- (211) Davis PL, Miron A, Andersen LM, Iglehart JD, Marks JR. Isolation and initial characterization of the BRCA2 promoter. *Oncogene* 1999 Oct 28;18(44):6000-12.
- (212) Wang J, Bian C, Li J, et al. Poly(ADP-ribose) polymerase-1 down-regulates BRCA2 expression through the BRCA2 promoter. *J Biol Chem* 2008 Dec 26;283(52):36249-56.
- (213) Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 2000 Dec 28;343(26):1910-6.
- (214) Migliazza A, Bosch F, Komatsu H, et al. Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood* 2001 Apr 1;97(7):2098-104.
- (215) Klein U, Lia M, Crespo M, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell* 2010 Jan 19;17(1):28-40.
- (216) Huang M, Kennedy R, Ali AM, et al. Human MutS and FANCM complexes function as redundant DNA damage sensors in the Fanconi Anemia pathway. *DNA Repair (Amst)* 2011 Dec 10;10(12):1203-12.

- (217) Horejsi Z, Collis SJ, Boulton SJ. FANCM-FAAP24 and HCLK2: roles in ATR signalling and the Fanconi anemia pathway. *Cell Cycle* 2009 Apr 15;8(8):1133-7.
- (218) Meetei AR, Medhurst AL, Ling C, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* 2005 Sep;37(9):958-63.
- (219) Wang X, Andreassen PR, D'Andrea AD. Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol Cell Biol* 2004 Jul;24(13):5850-62.
- (220) Xue Y, Li Y, Guo R, Ling C, Wang W. FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. *Hum Mol Genet* 2008 Jun 1;17(11):1641-52.
- (221) Collis SJ, Ciccia A, Deans AJ, et al. FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell* 2008 Nov 7;32(3):313-24.
- (222) Aronchik I, Bjeldanes LF, Firestone GL. Direct inhibition of elastase activity by indole-3-carbinol triggers a CD40-TRAF regulatory cascade that disrupts NF-kappaB transcriptional activity in human breast cancer cells. *Cancer Res* 2010 Jun 15;70(12):4961-71.
- (223) Zhou Y, Eppenberger-Castori S, Marx C, et al. Activation of nuclear factor-kappaB (NFkappaB) identifies a high-risk subset of hormone-dependent breast cancers. *Int J Biochem Cell Biol* 2005 May;37(5):1130-44.
- (224) Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer* 2010 Jun;10(6):389-402.
- (225) Landi D, Gemignani F, Naccarati A, et al. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis* 2008 Mar;29(3):579-84.
- (226) Mishra PJ, Bertino JR. MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine. *Pharmacogenomics* 2009 Mar;10(3):399-416.
- (227) Yu Z, Li Z, Jolicoeur N, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res* 2007;35(13):4535-41.
- (228) Gozgit JM, Pentecost BT, Marconi SA, et al. Use of an aggressive MCF-7 cell line variant, TMX2-28, to study cell invasion in breast cancer. *Mol Cancer Res* 2006 Dec;4(12):905-13.
- (229) Zhao T, Guan L, Yu Y, et al. Kindlin-2 promotes genome instability in breast cancer cells. *Cancer Lett* 2013 Apr 28;330(2):208-16.
- (230) Powers MA, Fay MM, Factor RE, Welm AL, Ullman KS. Protein arginine methyltransferase 5 accelerates tumor growth by arginine methylation of the tumor suppressor programmed cell death 4. *Cancer Res* 2011 Aug 15;71(16):5579-87.
- (231) Wen YH, Shi X, Chiriboga L, et al. Alterations in the expression of PDCD4 in ductal carcinoma of the breast. *Oncol Rep* 2007 Dec;18(6):1387-93.
- (232) Beroukhim R, Getz G, Nghiemphu L, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 2007 Dec 11;104(50):20007-12.

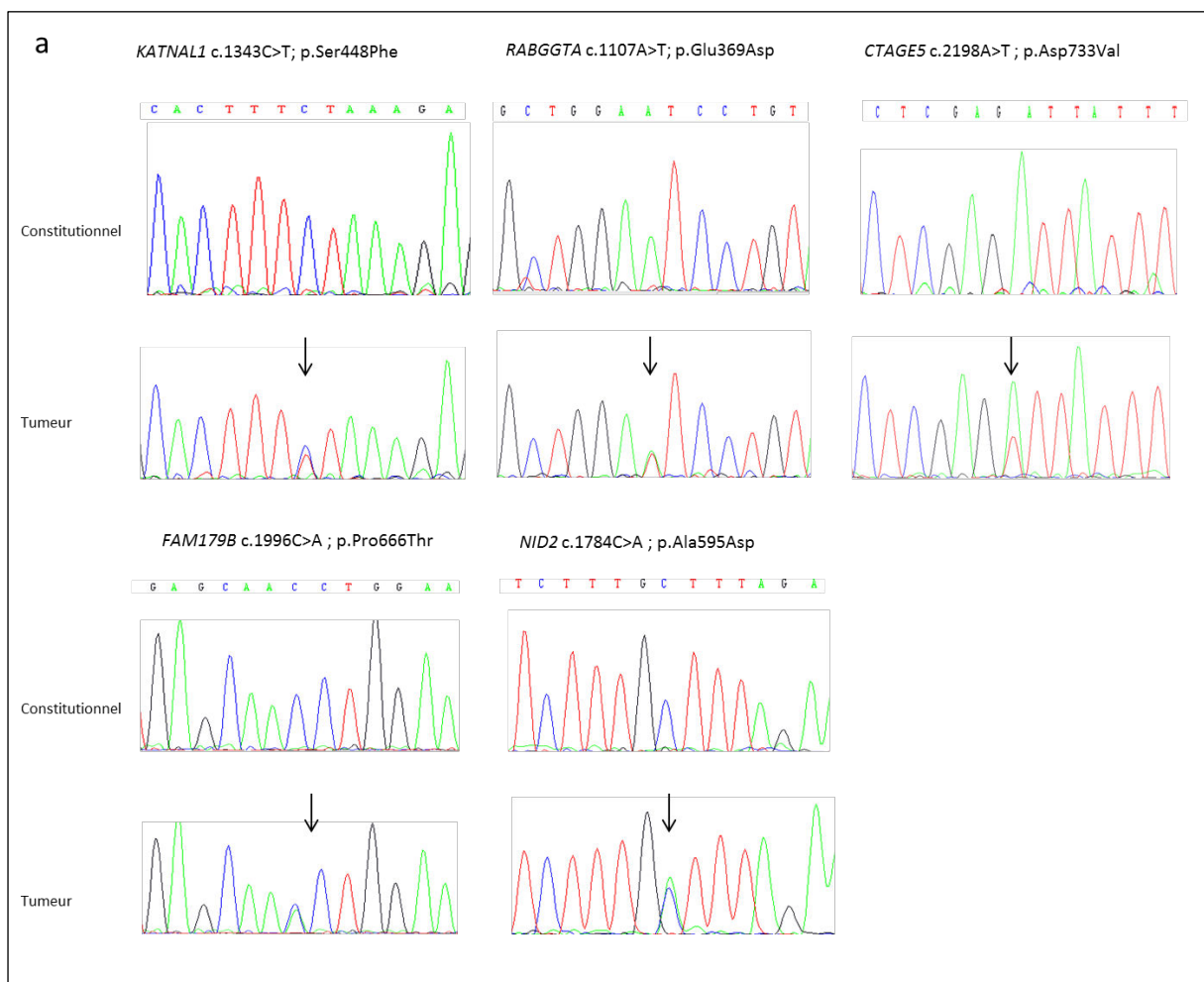
- (233) McCabe N, Turner NC, Lord CJ, et al. Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer Res* 2006 Aug 15;66(16):8109-15.
- (234) Cao L, Kim S, Xiao C, et al. ATM-Chk2-p53 activation prevents tumorigenesis at an expense of organ homeostasis upon Brca1 deficiency. *EMBO J* 2006 May 17;25(10):2167-77.
- (235) Schwartz M, Zlotorynski E, Goldberg M, et al. Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability. *Genes Dev* 2005 Nov 15;19(22):2715-26.
- (236) Huebner K, Croce CM. FRA3B and other common fragile sites: the weakest links. *Nat Rev Cancer* 2001 Dec;1(3):214-21.
- (237) Denison SR, Simper RK, Greenbaum IF. How common are common fragile sites in humans: interindividual variation in the distribution of aphidicolin-induced fragile sites. *Cytogenet Genome Res* 2003;101(1):8-16.
- (238) Mao G, Pan X, Gu L. Evidence that a mutation in the MLH1 3'-untranslated region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed leukemia. *J Biol Chem* 2008 Feb 8;283(6):3211-6.
- (239) Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med* 2000 Jun;6(6):637-41.
- (240) Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell* 2009 May;101(5):251-62.
- (241) Bansal V, Libiger O. A probabilistic method for the detection and genotyping of small indels from population-scale sequence data. *Bioinformatics* 2011 Aug 1;27(15):2047-53.
- (242) CASAVA v.1.8.2 User Guide. Illumina 2013 [cited 2013 Aug]; Available from: URL: [http://support.illumina.com/sequencing/sequencing\\_software/casava.ilmn](http://support.illumina.com/sequencing/sequencing_software/casava.ilmn)
- (243) Levin JZ, Berger MF, Adiconis X, et al. Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol* 2009;10(10):R115.

# **ANNEXES**

---

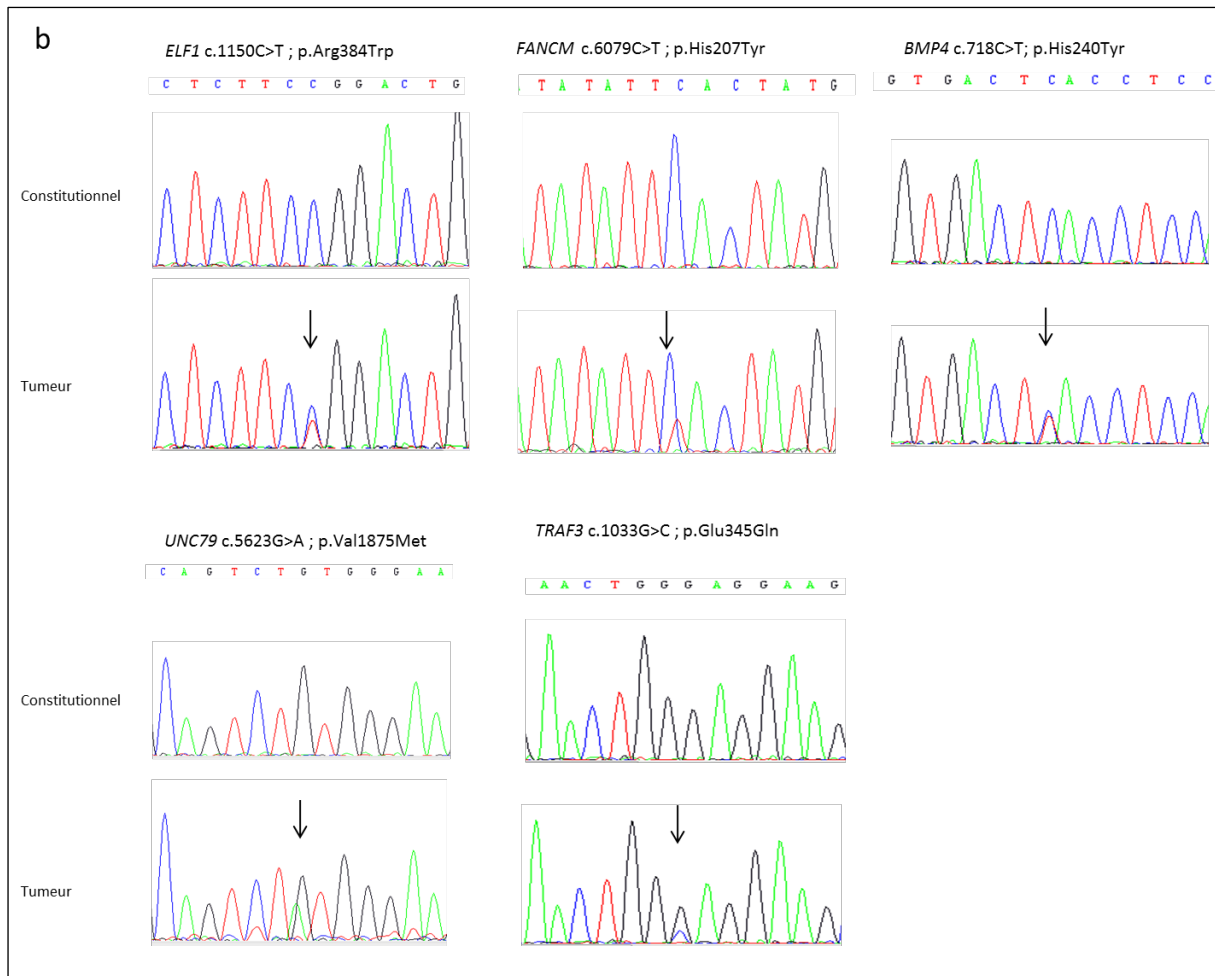
## 5. ANNEXES

**Annexe 1 : Electrophérogramme des 10 variants SNVs somatiques exoniques non iso-sémantiques séquencés sur la tumeur et l'ADN constitutionnel correspondant, (a) séquençage sur la tumeur 148 et l'ADN constitutionnel correspondant des variants des gènes *KATNAL1* c.1343C>T;p.Ser448Phe, *RABGGTA* c.1107A>T;p.Glu369Asp, *CTAGE5* c.2198A>T;p.Asp733Val, *FAM179B* c.1996C>A; p.Pro666Thr, *NID2* c.1784C>A;p.Ala595Asp ; (b) séquençage sur la tumeur 149 et l'ADN constitutionnel correspondant des variants des gènes: *ELF1* c.1150C>T;p.Arg384Trp, *FANCM* c.6079C>T;p.His2027Tyr, *BMP4* c.718C>T;p.His240Tyr, *UNC79*c.5623G>A;p.Val1875Met et *TRAF3* c.1033G>C;p.Glu345Gln.**

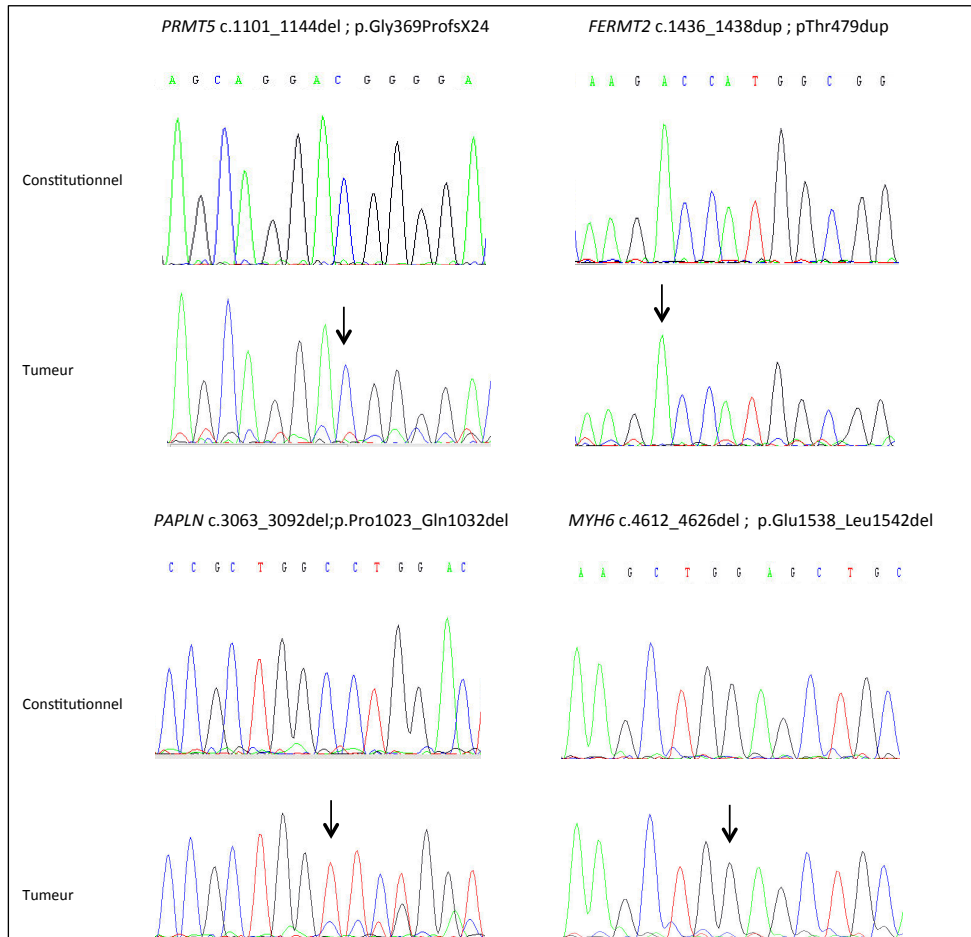




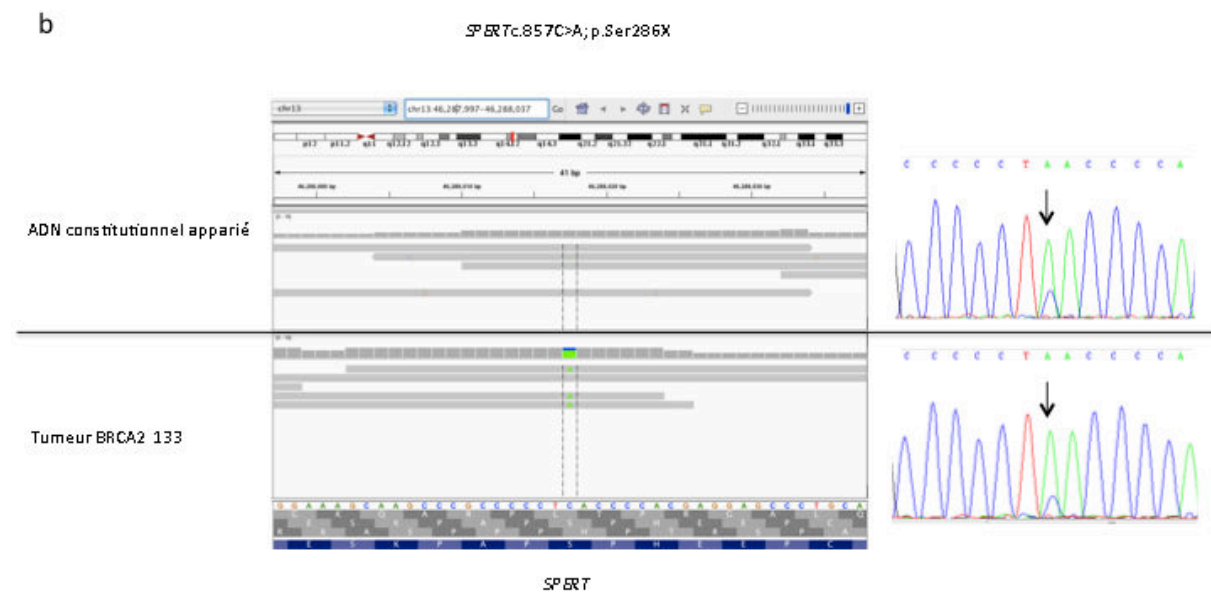
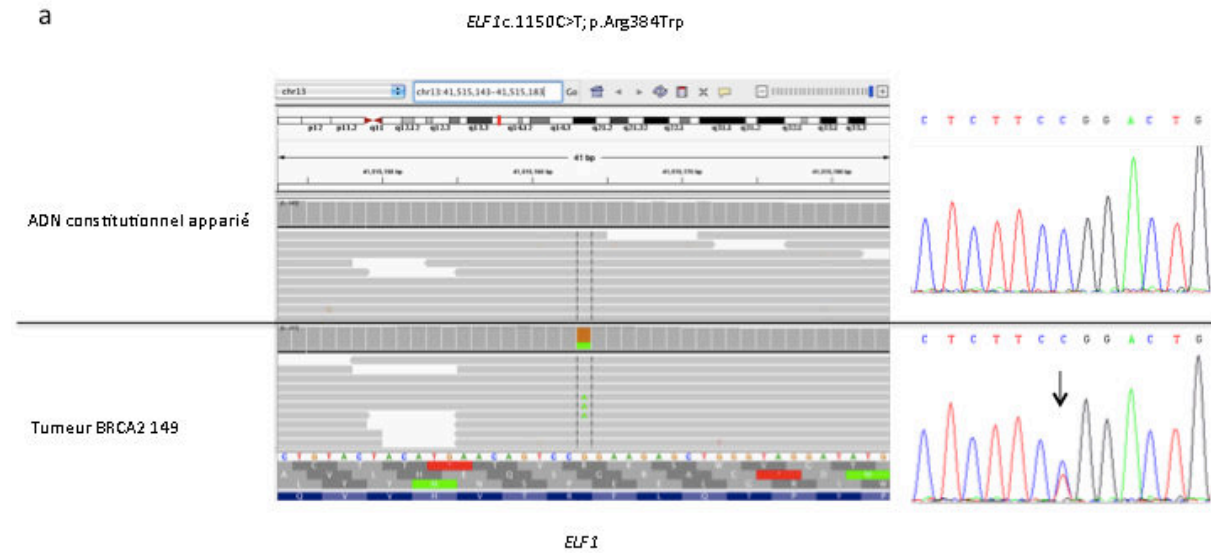
**Annexe 1 (suite): Electrophérogamme des 10 variants SNVs somatiques exoniques non iso-sémantiques séquencés sur la tumeur et l'ADN constitutif correspondant, (a) séquençage sur la tumeur 148 et l'ADN constitutif correspondant des variants des gènes *KATNAL1* c.1343C>T;p.Ser448Phe, *RABGGTA* c.1107A>T;p.Glu369Asp, *CTAGE5* c.2198A>T;p.Asp733Val, *FAM179B* c.1996C>A; p.Pro666Thr, *NID2* c.1784C>A;p.Ala595Asp ; (b) séquençage sur la tumeur 149 et l'ADN constitutif correspondant des variants des gènes: *ELF1* c.1150C>T;p.Arg384Trp, *FANCM* c.6079C>T;p.His207Tyr, *BMP4* c.718C>T;p.His240Tyr, *UNC79* c.5623G>A;p.Val1875Met et *TRAF3* c.1033G>C;p.Glu345Gln.**



**Annexe 2 : Electrophérogramme des 4 variants INDELs somatiques exoniques séquencés sur la tumeur et l'ADN constitutionnel correspondant (tumeur106 :PRMT5 c.1101\_1144del;p.Gly369ProfsX24 et FERMT2 c.1469\_14738dup;pThr479dup ; tumeur 148 : PAPLN c.3063\_3092del;p.Pro1023\_Gln1032del ;tumeur149:MYH6 c.4612\_4626del;p.Glu1538\_Leu1542del).**



**Annexe 3: Représentation de la visualisation des fichiers d'alignement .bam par l'outil IGV sur la tumeur et l'ADN constitutionnel correspondant. (a) Variant du gène *ELF1* confirmé en Sanger en somatique (tumeur 149) et absent sur le constitutionnel, (b) Variant du gène *SPERT* confirmé en Sanger en somatique (tumeur 133) et sur le constitutionnel (c) Variant du gène *PABPC3* non confirmé en Sanger en somatique (tumeur 149) et absent sur le constitutionnel (d) Variant du gène *ARL11* annoté somatique potentiel en raison d'une LOH dans la tumeur 148.**





## Annexe 4 : Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations

OPEN ACCESS Freely available online



# Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations

Audrey Rouault<sup>1</sup>, Guillaume Banneau<sup>1</sup>, Gaëtan MacGrogan<sup>1,2</sup>, Natalie Jones<sup>1</sup>, Nabila Elarouci<sup>3</sup>, Emmanuelle Barouk-Simonet<sup>4</sup>, Laurence Venat<sup>5</sup>, Isabelle Coupier<sup>6</sup>, Eric Letouzé<sup>3</sup>, Aurélien de Reyniès<sup>3</sup>, Françoise Bonnet<sup>1,4</sup>, Richard Iggo<sup>1\*</sup>, Nicolas Sévenet<sup>1,4</sup>, Michel Longy<sup>1,4</sup>

1 French National Institute of Health and Medical Research (INSERM) Unit 916, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France, 2 Pathology Department, Bergonié Cancer Institute, Bordeaux, France, 3 Cartes d'Identité des Tumeurs (CIT) Program, Ligue Nationale Contre Le Cancer, Paris, France, 4 Cancer Genetics Unit, Bergonié Cancer Institute, Bordeaux, France, 5 Department of Medical Oncology, Dupuytren University Hospital, Limoges, France, 6 Cancer Genetics Unit, Val d'Aurelle Regional Cancer Centre, Montpellier, France

# Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with BRCA2 Mutations

Audrey Rouault<sup>1</sup>, Guillaume Banneau<sup>1</sup>, Gaëtan MacGrogan<sup>1,2</sup>, Natalie Jones<sup>1</sup>, Nabila Elarouci<sup>3</sup>, Emmanuelle Barouk-Simonet<sup>4</sup>, Laurence Venat<sup>5</sup>, Isabelle Coupier<sup>6</sup>, Eric Letouzé<sup>3</sup>, Aurélien de Reyniès<sup>3</sup>, Françoise Bonnet<sup>1,4</sup>, Richard Iggo<sup>1\*</sup>, Nicolas Sévenet<sup>1,4</sup>, Michel Longy<sup>1,4</sup>

**1** French National Institute of Health and Medical Research (INSERM) Unit 916, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France, **2** Pathology Department, Bergonié Cancer Institute, Bordeaux, France, **3** Cartes d'Identité des Tumeurs (CIT) Program, Ligue Nationale Contre Le Cancer, Paris, France, **4** Cancer Genetics Unit, Bergonié Cancer Institute, Bordeaux, France, **5** Department of Medical Oncology, Dupuytren University Hospital, Limoges, France, **6** Cancer Genetics Unit, Val d'Aurelle Regional Cancer Centre, Montpellier, France

## Abstract

**Introduction:** Germline *BRCA1* or *BRCA2* mutations account for 20–30% of familial clustering of breast cancer. The main indication for *BRCA2* screening is currently the family history but the yield of mutations identified in patients selected this way is low.

**Methods:** To develop more efficient approaches to screening we have compared the gene expression and genomic profiles of *BRCA2*-mutant breast tumors with those of breast tumors lacking *BRCA1* or *BRCA2* mutations.

**Results:** We identified a group of 66 genes showing differential expression in our training set of 7 *BRCA2*-mutant tumors and in an independent validation set of 19 *BRCA2*-mutant tumors. The differentially expressed genes include a prominent cluster of genes from chromosomes 13 and 14 whose expression is reduced. Gene set enrichment analysis confirmed that genes in specific bands on 13q and 14q showed significantly reduced expression, suggesting that the affected bands may be preferentially deleted in *BRCA2*-mutant tumors. Genomic profiling showed that the *BRCA2*-mutant tumors indeed harbor deletions on chromosomes 13q and 14q. To exploit this information we have created a simple fluorescence in situ hybridization (FISH) test and shown that it detects tumors with deletions on chromosomes 13q and 14q.

**Conclusion:** Together with previous reports, this establishes that deletions on chromosomes 13q and 14q are a hallmark of *BRCA2*-mutant tumors. We propose that FISH to detect these deletions would be an efficient and cost-effective first screening step to identify potential *BRCA2*-mutation carriers among breast cancer patients without a family history of breast cancer.

**Citation:** Rouault A, Banneau G, MacGrogan G, Jones N, Elarouci N, et al. (2012) Deletion of Chromosomes 13q and 14q Is a Common Feature of Tumors with *BRCA2* Mutations. PLoS ONE 7(12): e52079. doi:10.1371/journal.pone.0052079

**Editor:** Sandra Orsulic, Cedars-Sinai Medical Center, United States of America

**Received:** August 22, 2012; **Accepted:** November 8, 2012; **Published:** December 21, 2012

**Copyright:** © 2012 Rouault et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work was funded by grants from the Charente-Maritime and Pyrenees-Atlantiques committees of the French Cancer League and the Bergerac Lions Club to M.L.; the French National Research Agency (ANR) to R.I.; and a French National Cancer League program grant to M.L. and R.I. This work is part of the national program Cartes d'Identité des Tumeurs (CIT) [http://cit.ligue-cancer.net/] funded and developed by the Ligue nationale contre le cancer. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: r.iggo@bordeaux.unicancer.fr

## Introduction

Germline mutations in pathways critical for maintenance of genomic integrity confer an increased risk of developing breast cancer [1]. Inherited mutations in two genes, breast cancer 1 (*BRCA1*) and *BRCA2*, are associated with a particularly striking increase in breast cancer risk [2]. Consistent with the Knudson two-hit model, both alleles of *BRCA1* and *BRCA2* are inactivated in tumors, indicating that the genes behave like classic tumor suppressor genes [3]. Their gene products are implicated in the repair of DNA double-strand breaks [4]: *BRCA1* is required for recruitment of repair proteins to sites of breakage [5], whereas *BRCA2* nucleates RAD51 filament assembly on single-stranded DNA exposed by resection from the break [6]. Loss of these functions leads to genomic instability [7].

The criteria used to select patients for *BRCA2* screening are essentially based on the family history. Unfortunately, this approach is wasteful of resources because relatively few familial clusters are caused by germline *BRCA2* mutations [8]. This approach also overlooks patients with no overt family history of breast or ovarian cancer who may nevertheless have *BRCA2* mutations. Despite numerous efforts, no specific clinical or pathological features have been identified that permit easy identification of *BRCA2*-associated tumors.

The role *BRCA2* plays in repair of double strand breaks by homologous recombination might be expected to give a characteristic pattern of genomic instability but no genomic features have yet been described that can be used to identify these tumors. Gene expression profiling typically places the tumors in the luminal B, high proliferation, estrogen receptor (ER) positive group of the

**Table 1.** Characteristics of patients and tumors.

ID	Tumor set	BRCA status	Sex	Age at surgery (year)	Tumor size (mm)	Tumor cells (%)	Histologic grade	ER	PR	ERBB2
52	Training	BRCA2	F	35	17	92	3	++	–	–
86	Training	BRCA2	F	46	16	90	3	+++	+	–
106	Training	BRCA2	F	57	22	85	3	+++	+	–
133	Training	BRCA2	F	40	15	75	2	+	+	–
144	Training	BRCA2	F	40	12	55	2	++	–	+
146	Training	BRCA2	F	64	25	80	3	–	–	+
148	Training	BRCA2	F	62	25	90	3	++	–	++
8	Training	BRCAX	F	51	18	90	3	–	–	–
9	Training	BRCAX	F	51	25	95	3	++	++	–
11	Training	BRCAX	F	56	40	78	2	++	+++	–
14	Training	BRCAX	F	45	12	90	2	nd	+++	–
16	Training	BRCAX	F	50	27	95	3	+++	+++	–
22	Training	BRCAX	F	64	18	90	2	+++	+	–
24	Training	BRCAX	F	35	12	70	1	++	–	–
25	Training	BRCAX	F	37	12	92	2	++	+	–
33	Training	BRCAX	F	42	35	73	1	++	–	–
37	Training	BRCAX	F	45	20	92	2	+++	+++	–
38	Training	BRCAX	F	64	13	90	3	+++	–	–
40	Training	BRCAX	F	41	12	95	2	+	++	–
41	Training	BRCAX	F	38	21	92	3	++	+	–
46	Training	BRCAX	F	60	38	90	2	–	–	–
66	Training	BRCAX	F	73	12	90	2	+++	+++	–
75	Training	BRCAX	F	58	14	80	2	–	–	+++
79	Training	BRCAX	F	42	11	90	3	++	++	–
81	Training	BRCAX	F	46	28	80	2	+	++	–
82	Training	BRCAX	F	50	9	85	1	++	++	–
84	Training	BRCAX	F	47	27	92	3	++	+	+
85	Training	BRCAX	F	64	15	90	1	–	+++	–
93	Training	BRCAX	F	44	18	85	2	+++	+++	–
107	Training	BRCAX	F	69	40	80	3	+++	+	–
111	Training	BRCAX	F	73	15	80	1	+++	+	–
3	Validation	BRCAX	F	36	18	95	3	–	+	+++
15	Validation	BRCAX	F	42	15	95	3	–	–	–
17	Validation	BRCAX	F	76	3	95	1	+++	+	–
30	Validation	BRCAX	F	51	nd	95	3	+++	–	nd
48	Validation	BRCAX	F	54	20	90	1	++	++	–
49	Validation	BRCAX	F	49	35	66	2	+++	++	–
65	Validation	BRCAX	F	46	37	95	3	–	–	–
71	Validation	BRCAX	F	43	21	73	2	++	+++	+++
83	Validation	BRCAX	F	50	18	50	2	++	–	–
89	Validation	BRCAX	F	30	30	82	nd	++	+++	–
96	Validation	BRCAX	F	41	25	85	3	–	–	+++
99	Validation	BRCAX	M	63	21	90	1	+++	++	–
43	Genomic	BRCA2	F	38	12	90	2	++	–	–
149	Genomic	BRCA2	F	76	70	60	2	+++	–	–

Footnote. Tumor set: Training set, tumors used to create the gene expression signature; Validation set, BRCAX tumors from Bergonie Cancer Institute; Genomic set, tumors only used for CGH and SNP analysis. nd, not determined. There was no statistically significant difference ( $p > 0.05$ , Fisher test) between the BRCA2 and BRCAX groups for the following comparisons: age at surgery  $<$  vs  $\geq 49$  years (median age); tumor size  $<$  vs  $\geq 18$  mm (median tumor size); tumor cell content  $<$  vs  $\geq 90\%$  (median tumor cell content); +++ vs other ER status; – vs other PR status; – vs other ERBB2 status.

doi:10.1371/journal.pone.0052079.t001

Stanford classification but this is not specific enough to be useful clinically to identify tumors with *BRCA2* mutations [9].

In this study, we have used gene expression and genomic data to identify specific molecular features that distinguish tumors with *BRCA2* mutations from tumors with other breast cancer predisposition mutations. Based on these results we have developed a fluorescent *in-situ* hybridization (FISH) test that can be used to screen for tumors with an increased risk of containing *BRCA2* mutations.

## Methods

### Patients and Samples

All samples were from the Bergonie Cancer Institute, Bordeaux, except for sample 144 from the Val d'Aurelle Regional Cancer Center, Montpellier; samples 146 and 148 from the Dupuytren Hospital, Limoges; and the *BRCA2* tumors in the validation set from the Curie Institute. The microarray data for the validation set were generously provided by the Translational Research Unit at the Curie Institute, Paris. The control group contained *BRCA2* tumors, defined as tumors lacking known *BRCA1/2* mutations from families with either i) at least three breast cancer-affected first or second-degree relatives; or ii) breast cancer before age 42 or ovarian cancer in two first-degree relatives or two second-degree relatives via a male. 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 authorisation number AC-2008-812, which was approved by the Comité de Protection des Personnes. The *BRCA1* and *BRCA2* mutation search was made after patients gave signed informed consent in the context of a medical genetic diagnosis of suspected breast cancer predisposition, in compliance with the French law on genetic testing (law number 94-654).

### Tumor and Mutation Characterization

Clinical, pathological and genetic data for each case are listed in Table 1. Immunohistochemistry for ER, progesterone receptor (PR) and HER2 (ERBB2) were performed as previously described [10]. HER2 expression was scored according to the Herceptest system. ER and PR were scored by multiplying the percentage of positive cells by the intensity (score 0–20: –; score 21–100: +; score 101–200: ++; score 201–300: +++). Screening for germline mutations was performed on leucocyte DNA as previously described [10].

### Gene Expression and Genomic Chip Hybridization

RNA was extracted from the tumors as described [10] and hybridized to Affymetrix U133 Plus 2.0 genechip microarrays by the Genopole Alsace-Lorraine genomics platform, except for the validation set which was hybridized by the Curie Institute genomics platform. DNA was extracted from the tumors and hybridized to Integrachip V7 bacterial artificial chromosome (BAC) arrays as described [10]. SNP array profiling was performed on Illumina Human610-Quad v1.0 BeadChips (Illumina, Inc., San Diego, CA) by Integragen (Evry, France). The gene expression and genomic data are available in Array Express under accession numbers E-TABM-854, E-MEXP-3688, E-MEXP-3690 and in GEO under accession number GSE39710.

### Data Processing and Statistical Analyses

Given the rarity of the tumors, it was not possible to avoid processing the tumors in batches; the hybridization dates for the Affymetrix chips are given in the CEL files. The 12 *BRCA2*

controls for the validation set were chosen because they showed the smallest batch effect relative to the Curie Institute tumors. The 12 *BRCA2* tumors in the validation set were separate from the 24 *BRCA2* tumors in the training set. The gene expression data were normalized with the RMA algorithm in R version 2.13.1 [11–13]. To eliminate redundant genes sharing a gene symbol, the most variable probeset was selected based on the standard deviation across the entire dataset. Differentially expressed genes were identified by moderated t-test in limma [14] (an R script for the expression analysis is available on request). The 66 *BRCA2* gene signature genes were combined to make a *BRCA2* score by summing the mean-centered expression values weighted by the t values from limma. Gene Set Enrichment Analysis (GSEA) was performed with Broad Institute java software [15,16]: the expression dataset was ranked by t-statistic in limma, then enrichment was scored by GSEA for chromosome bands using the MSigDB positional gene sets [15,16]. Centroid-linkage hierarchical clustering was performed in Cluster 3.0 and visualized in TreeView [17]. Array CGH data was normalized with CAPweb software [18] and genomic alterations were visualized with VAMP software using the same thresholds as previously described [10]. SNP data were normalized with Illumina Genome Studio Software v2010.1 using Genotyping module (v1.6.3) and Illumina Genome Viewer module (v1.6.1) to obtain the B Allele Frequency (BAF) for each SNP.

### Fluorescence *In Situ* Hybridization

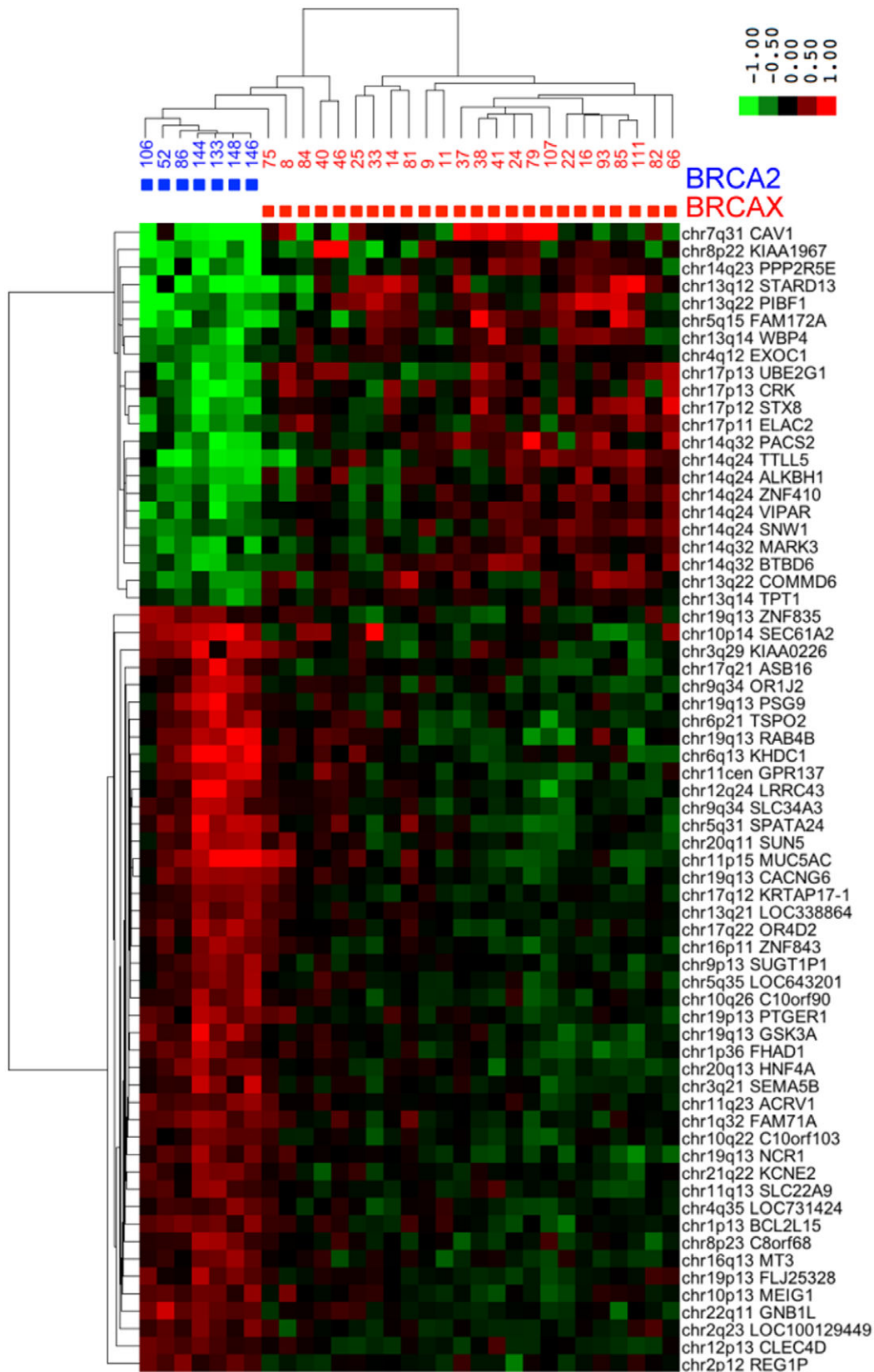
To detect deletions on chromosomes 13 and 14, FISH was performed with four BAC probes supplied by BlueGnome (Cambridge, UK). Two clones labeled with SpectrumGreen were used to detect the pericentromeric regions of chromosomes 13 and 14: RP11-408E5 on 13q12.11 (hg19 chr13:19700993–19850551); and RP11-98N22 on 14q11.2 (hg19 chr14:20500968–20660726). Two clones labeled with SpectrumOrange (giving red spots in the figures) were used to detect the deletions on chromosomes 13 and 14: RP11-71C5 on 13q14.11 (hg19 chr13:44921196–45086777) and RP11-242P2 on 14q31.1 (hg19 chr14:80030106–80193689). Nuclei obtained by touch imprints were fixed in 3:1 methanol: acetic acid, washed and dried. The BAC probes were mixed, 5  $\mu$ l of hybridization mix was added per slide, and a coverslip was glued in place to create a hybridization chamber. The sections were denatured at 75°C for 5 minutes and hybridized at 37°C overnight. Stringent washes were performed at 65°C for 10 minutes, then the sections were dehydrated in ethanol and mounted. Images were acquired with a Zeiss Axio Imager Z2 microscope (Gottingen, Germany). The number of red and green spots per nucleus was scored in morphologically intact and non-overlapping nuclei. Deletions were reported when  $\geq 50\%$  of nuclei with the modal number of green spots contained fewer red spots or when they contained single green and red spots.

## Results

### Identification of Genes Differentially Expressed in *BRCA2*-mutant Tumors

To gain insight into the biology of *BRCA2*-mutant breast tumors, we performed a supervised analysis looking for genes differentially expressed in *BRCA2*-mutant and control tumors. All of the tumors came from patients with a familial clustering of breast cancer potentially caused by germline mutation of a breast cancer predisposition gene. The *BRCA2*-mutant group included 7 tumors from patients with known germline *BRCA2* mutations. The control group (“*BRCA2*”) contained 24 patients without mutations in *BRCA1* or *BRCA2* identifiable by conventional screening. RNA





**Figure 1. Unsupervised hierarchical clustering of the 66 *BRCA2* signature genes in the training set.** There are seven *BRCA2*-mutant tumors and 24 *BRCA2* tumors (tumors from patients lacking known *BRCA1/2* mutations but with a familial history of breast cancer). The upper left quadrant contains many genes on 13q and 14q that show reduced expression in *BRCA2* tumors.  
doi:10.1371/journal.pone.0052079.g001

**Table 2.** BRCA2 signature genes.

Affymetrix ID	Gene Symbol	Gene Description	Band	t	p
222127_s_at	EXOC1	exocyst complex component 1	4q12	-7.05	0.0011
223564_s_at	GNB1L	G protein beta polypeptide 1-like	22q11	6.85	0.0011
632_at	GSK3A	glycogen synthase kinase 3 alpha	19q13	6.42	0.0025
1555377_at	OR4D2	olfactory receptor, family 4, subfamily D, member 2	17q22	6.13	0.0030
208429_x_at	HNF4A	hepatocyte nuclear factor 4, alpha	20q13	6.12	0.0030
207973_x_at	ACRV1	acrosomal vesicle protein 1	11q23	6.21	0.0030
218431_at	C14orf133	VPS33B interacting protein	14q24	-6.01	0.0034
1552510_at	SLC34A3	solute carrier family 34 (sodium phosphate), member 3	9q34	5.9	0.0041
204690_at	STX8	syntaxin 8	17p12	-5.84	0.0044
227630_at	PPP2R5E	protein phosphatase 2, regulatory subunit B', epsilon	14q23	-5.7	0.0047
205621_at	ALKBH1	alkB, alkylation repair homolog 1 (E. coli)	14q24	-5.69	0.0047
202569_s_at	MARK3	MAP/microtubule affinity-regulating kinase 3	14q32	-5.74	0.0047
216520_s_at	TPT1	tumor protein, translationally-controlled 1	13q14	-5.71	0.0047
230055_at	KHDC1	KH homology domain containing 1	6q13	5.6	0.0048
221966_at	GPR137	G protein-coupled receptor 137	11cen	5.62	0.0048
207733_x_at	PSG9	pregnancy specific beta-1-glycoprotein 9	19q13	5.59	0.0048
1555614_at	SUGT1P1	suppressor of G2 allele of SKP1 (S. cerevisiae) pseudogene 1	9p13	5.57	0.0048
1552772_at	CLEC4D	C-type lectin domain family 4, member D	12p13	5.57	0.0048
203598_s_at	WBP4	WW domain binding protein 4 (formin binding protein 21)	13q14	-5.51	0.0048
1563639_a_at	FHAD1	forkhead-associated (FHA) phosphopeptide binding domain 1	1p36	5.54	0.0048
234680_at	KRTAP17-1	keratin associated protein 17-1	17q12	5.52	0.0048
1562657_a_at	C10orf90	chromosome 10 open reading frame 90	10q26	5.45	0.0055
236979_at	BCL2L15	BCL2-like 15	1p13	5.39	0.0061
221095_s_at	KCNE2	potassium voltage-gated channel, Isk-related family, member 2	21q22	5.4	0.0061
213239_at	PIBF1	progesterone immunomodulatory binding factor 1	13q22	-5.36	0.0063
1567257_at	OR1J2	olfactory receptor, family 1, subfamily J, member 2	9q34	5.34	0.0064
225389_at	BTBD6	BTB (POZ) domain containing 6	14q32	-5.31	0.0066
207778_at	REG1P	regenerating islet-derived 1 pseudogene	2p12	5.3	0.0066
226005_at	UBE2G1	ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, yeast)	17p13	-5.25	0.0070
215424_s_at	SNW1	SNW domain containing 1	14q24	-5.23	0.0070
1564112_at	FAM71A	Family with sequence similarity 71, member A	1q32	5.25	0.0070
237980_at	LINC00347	hypothetical LOC338864	13q21	5.24	0.0070
213103_at	STARD13	StAR-related lipid transfer (START) domain containing 13	13q12	-5.18	0.0071
237257_at	RAB4B	RAB4B, member RAS oncogene family	19q13	5.19	0.0071
201767_s_at	ELAC2	elaC homolog 2 (E. coli)	17p11	-5.2	0.0071
209944_at	ZNF410	zinc finger protein 410	14q24	-5.16	0.0071
1558641_at	SPATA24	spermatogenesis associated 24	5q31	5.2	0.0071
212735_at	KIAA0226	Beclin-1 associated RUN domain containing protein	3q29	5.17	0.0071
215449_at	TSPO2	translocator protein 2	6p21	5.15	0.0071
1553253_at	ASB16	ankyrin repeat and SOCS box-containing 16	17q21	5.14	0.0071
231625_at	SLC22A9	solute carrier family 22 member 9	11q13	5.2	0.0071
225312_at	COMMD6	COMM domain containing 6	13q22	-5.12	0.0074
217187_at	MUC5AC	mucin 5AC, oligomeric mucus	11p15	5.1	0.0077
1553728_at	LRRC43	leucine rich repeat containing 43	12q24	5.07	0.0079
1552863_a_at	CACNG6	calcium channel, voltage-dependent, gamma subunit 6	19q13	5.07	0.0079
217095_x_at	NCR1	natural cytotoxicity triggering receptor 1	19q13	5.06	0.0079
223610_at	SEMA5B	semaphorin 5b	3q21	5.06	0.0079
203065_s_at	CAV1	caveolin 1, caveolae protein, 22 kDa	7q31	-5.03	0.0080
202226_s_at	CRK	v-crk sarcoma virus CT10 oncogene homolog (avian)	17p13	-5.04	0.0080

**Table 2.** Cont.

Affymetrix ID	Gene Symbol	Gene Description	Band	t	p
235416_at	LOC643201	centrosomal protein 192 kDa pseudogene	5q35	5.03	0.0080
1557827_at	C10orf103	chromosome 10 open reading frame 103	10q22	5.03	0.0080
225187_at	KIAA1967	DBC1 deleted in breast cancer 1	8p22	-4.98	0.0082
212936_at	FAM172A	family with sequence similarity 172, member A	5q15	-4.99	0.0082
215898_at	TTL5	tubulin tyrosine ligase-like family, member 5	14q24	-4.98	0.0082
212778_at	PACS2	phosphofurin acidic cluster sorting protein 2	14q32	-5	0.0082
1562914_a_at	FLJ25328	hypothetical LOC148231	19p13	5	0.0082
215826_x_at	ZNF835	zinc finger protein 835	19q13	4.97	0.0084
238158_at	MEIG1	meiosis expressed gene 1 homolog (mouse)	10p13	4.97	0.0084
219499_at	SEC61A2	Sec61 alpha 2 subunit ( <i>S. cerevisiae</i> )	10p14	4.94	0.0087
207650_x_at	PTGER1	prostaglandin E receptor 1 (subtype EP1), 42 kDa	19p13	4.94	0.0087
237188_x_at	SUN5	Sad1 and UNC84 domain containing 5	20q11	4.92	0.0091
1557679_at	C8orf68	chromosome 8 open reading frame 68	8p23	4.91	0.0092
224256_at	LOC100129449	PRO2055	2q23	4.89	0.0095
1564362_x_at	ZNF843	zinc finger protein 843	16p11	4.88	0.0097
205970_at	MT3	metallothionein 3	16q13	4.87	0.0098
1569095_at	LOC731424	hypothetical LOC731424	4q35	4.87	0.0098

Footnote. t: moderated t-statistic for 66 genes that best discriminate between *BRCA2* and *BRCA1* tumors. p: p-value after Benjamini Hochberg correction (all genes had an unadjusted p-value <0.0001).

doi:10.1371/journal.pone.0052079.t002

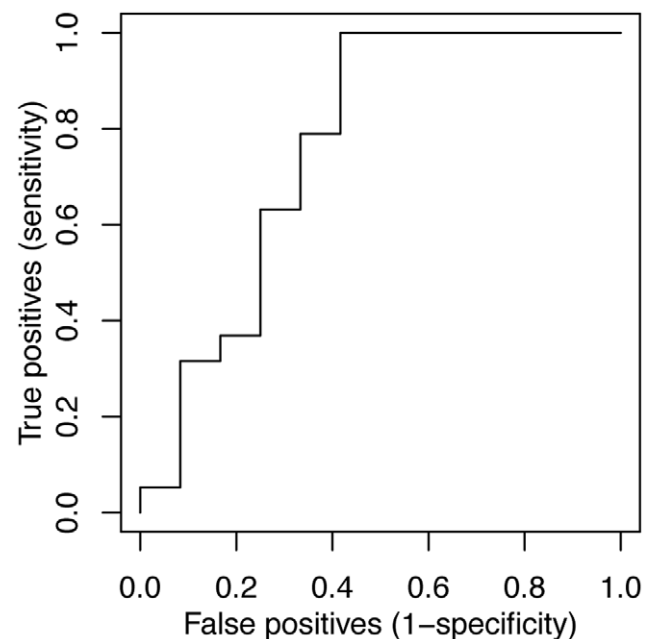
from these 31 tumors was tested on Affymetrix gene expression chips. Sixty-six genes were differentially expressed in the *BRCA2* and *BRCA1* groups at a false discovery rate <0.01 after Benjamini Hochberg correction for multiple testing (Table 2). Hierarchical clustering confirmed, as expected, that the differentially expressed genes cleanly split the tumors into two groups (Figure 1). The *BRCA2* group in the heatmap contains five *BRCA1* tumors that may represent tumors whose *BRCA2* mutations were missed by screening or tumors that phenocopy *BRCA2*.

### Validation of a Putative *BRCA2* Signature

We combined the differentially expressed genes in Table 2 to make a potential *BRCA2* gene expression signature. Receiver operating characteristic (ROC) analysis showed that the area under the curve (AUC) for classification of the training set was 1.0 with the *BRCA2* signature genes, indicating perfect classification of the tumors. This is not surprising given the small size of the dataset. To test for overfitting, we analyzed an independent validation set of 19 *BRCA2*-mutant tumors from the Curie Institute genetics clinic and 12 *BRCA1* from the Bergonie Cancer Institute. Given the rarity of the disease it is unfortunately difficult to avoid batch effects that might confound the result. Nevertheless, the AUC of the ROC curve was 0.76 in the validation set (Figure 2), indicating that the *BRCA2* signature was able to classify *BRCA2*-mutant tumors reasonably well. Hierarchical clustering confirmed that the *BRCA2* signature genes were differentially expressed in the validation set (Figure 3). While this suggests that the *BRCA2* signature has discriminant value in our tumors and in the validation set from the Curie Institute we note that this is not generally the case because the signature does not identify *BRCA2*-mutant tumors in some published datasets. For example, the AUC in the Waddell dataset [19] was 0.64, perhaps because of differences in the technology or in the populations studied. We conclude that the *BRCA2* signature may have discriminant value in tumors processed according to our protocol.

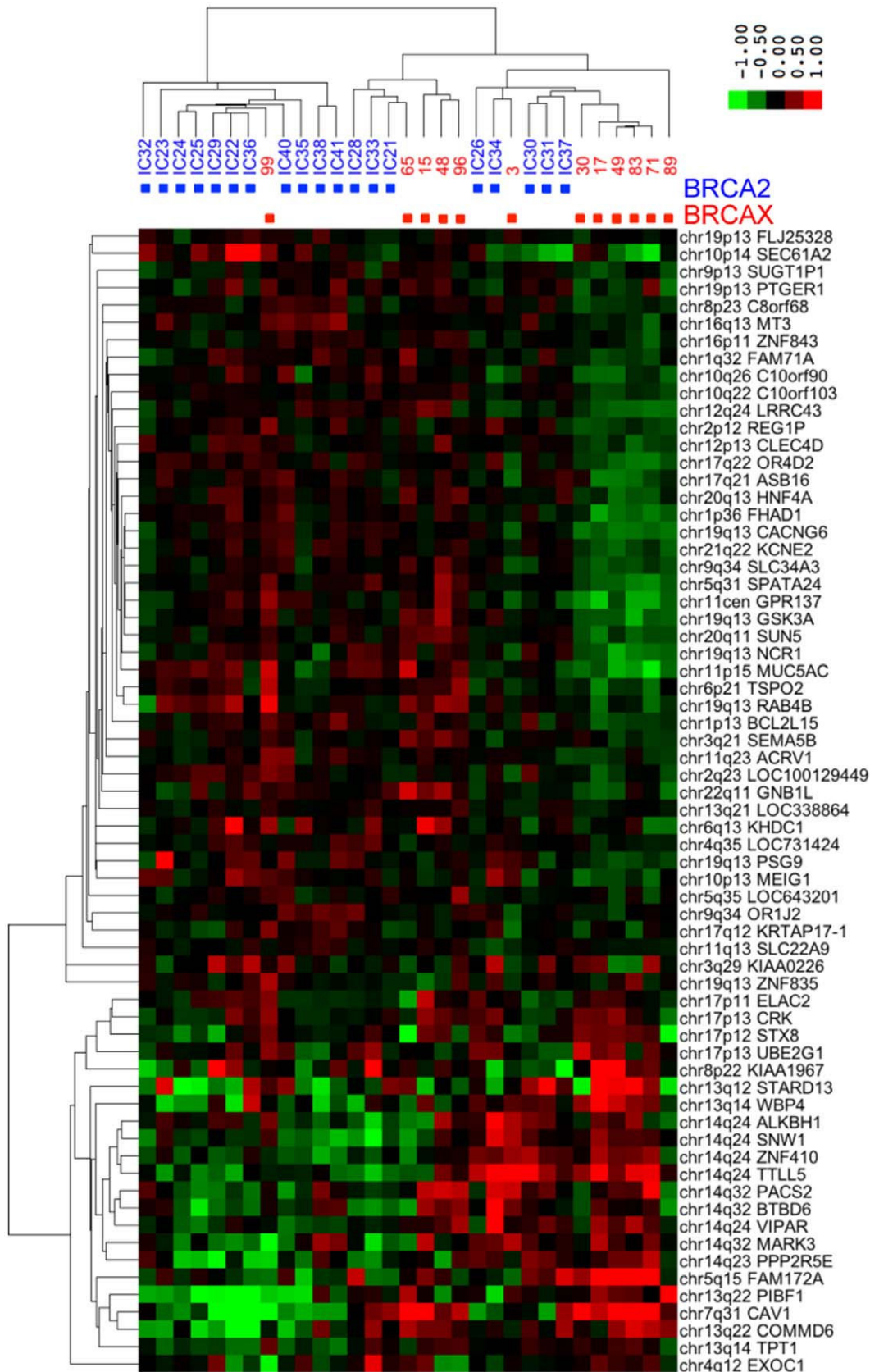
### Gene Set Enrichment Analysis (GSEA) Reveals the Mechanism Behind the *BRCA2* Signature

The striking feature of the heatmap in Figure 1 is the cluster of 22 genes showing reduced expression in *BRCA2*-mutant tumors.



**Figure 2. ROC analysis of the *BRCA2* signature in the validation set.** Each tumor was given a score that was a weighted sum of the mean centered gene expression levels for each gene in the signature. The validation set contained 19 *BRCA2* and 12 *BRCA1* tumors. The AUC was 0.76.

doi:10.1371/journal.pone.0052079.g002



**Figure 3. Unsupervised hierarchical clustering of the 66 *BRCA2* signature genes in the validation set.** There are 19 *BRCA2*-mutant tumors and 12 *BRCA2* tumors. The lower left quadrant contains many genes on 13q and 14q that show reduced expression in *BRCA2* tumors. doi:10.1371/journal.pone.0052079.g003

**Table 3. GSEA for loss of chromosomal bands.**

Band	Genes	ES	NES
13q14	67	-0.63	-2.75
14q31	22	-0.81	-2.71
13q13	22	-0.74	-2.45
14q24	77	-0.54	-2.43
17p13	185	-0.44	-2.3
14q32	105	-0.48	-2.28
10q26	72	-0.51	-2.27
4p16	91	-0.49	-2.25

Footnote. The genes column shows the number of genes used to score the band. The nominal, FDR and FWER p-values were all <0.001. ES, enrichment score; NES normalized enrichment score.

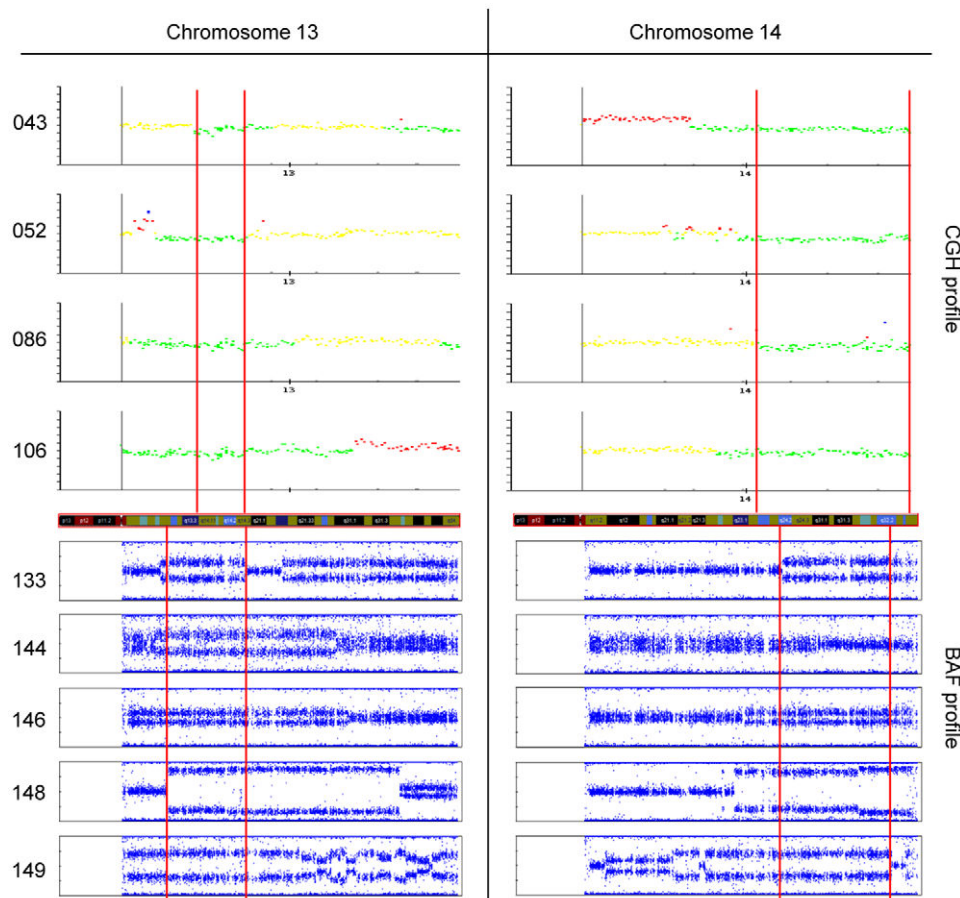
doi:10.1371/journal.pone.0052079.t003

These genes show a correlation of 0.90 in the heatmap. To exclude fortuitous hybridization as an explanation for this strong clustering we verified that the probe sequences were different and that they were labeled by Affymetrix as valid, non-cross-hybridizing probes for the indicated genes. Fourteen of the 22 *BRCA2* signature genes showing reduced expression are from

chromosomes 13 and 14. To determine whether this was due to chance, we ranked the dataset by moderated t statistic (*BRCA2* vs control), then performed GSEA with gene sets derived from individual chromosomal bands. The bands most frequently lost are shown in Table 3. The enrichment for bands on 13q and 14q was highly significant ( $p < 0.001$  for the family-wise error rate, the most stringent criterion in the Broad Institute implementation of GSEA). The most likely explanation for underexpressed genes to be derived from specific chromosomal bands is deletion of those bands in the corresponding tumors.

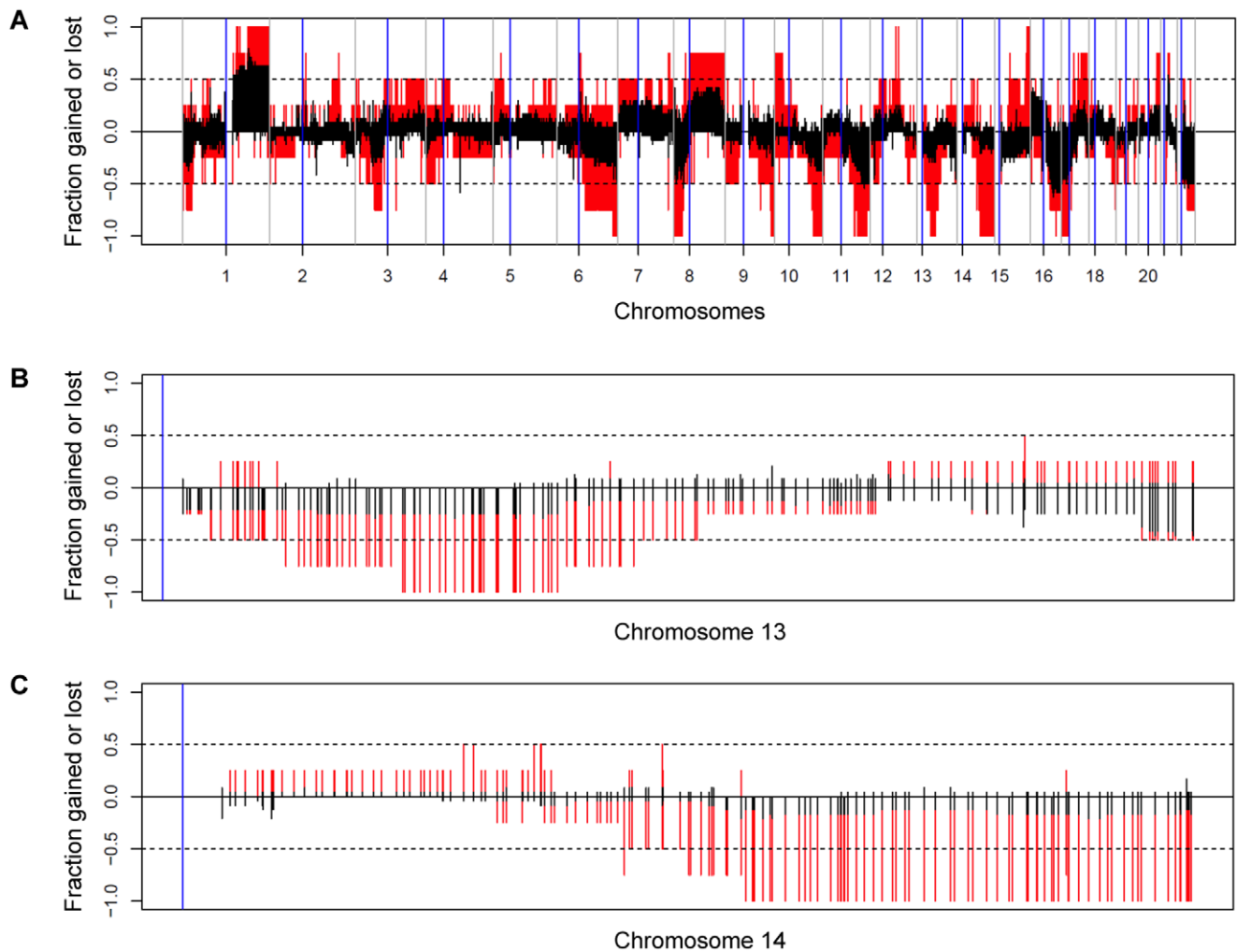
#### CGH and SNP Analysis of *BRCA2*-mutant Tumors

To test directly for loss of the regions containing the *BRCA2* signature genes we measured DNA copy number on CGH and SNP chips. The resulting CGH and SNP profiles confirmed that the incriminated regions are indeed deleted in the *BRCA2*-mutant tumors (Figure 4). The common region of overlap of the deletions extends from 13q13.3 to 13q14.3 and from 14q24.2 to 14q32.2. The cumulative rates of gain and loss for the *BRCA2* and *BRCA1* tumors are shown in Figure 5. This shows that the long arms of both chromosomes 13 and 14 contain large regions that are preferentially deleted in the *BRCA2*-mutant tumors. We conclude that the *BRCA2* signature genes are differentially expressed because they are deleted in the *BRCA2* tumors.



**Figure 4. Genomic profiles in the training set.** Upper panels: BAC-CGH profiles of *BRCA2*-mutant tumors showing gains in red, losses in green and modal copy number in yellow. Lower panels: BAF profiles of *BRCA2*-mutant tumors on Illumina SNP arrays. The boundaries of the common regions of deletion on chromosomes 13 and 14 are marked by vertical red lines.

doi:10.1371/journal.pone.0052079.g004



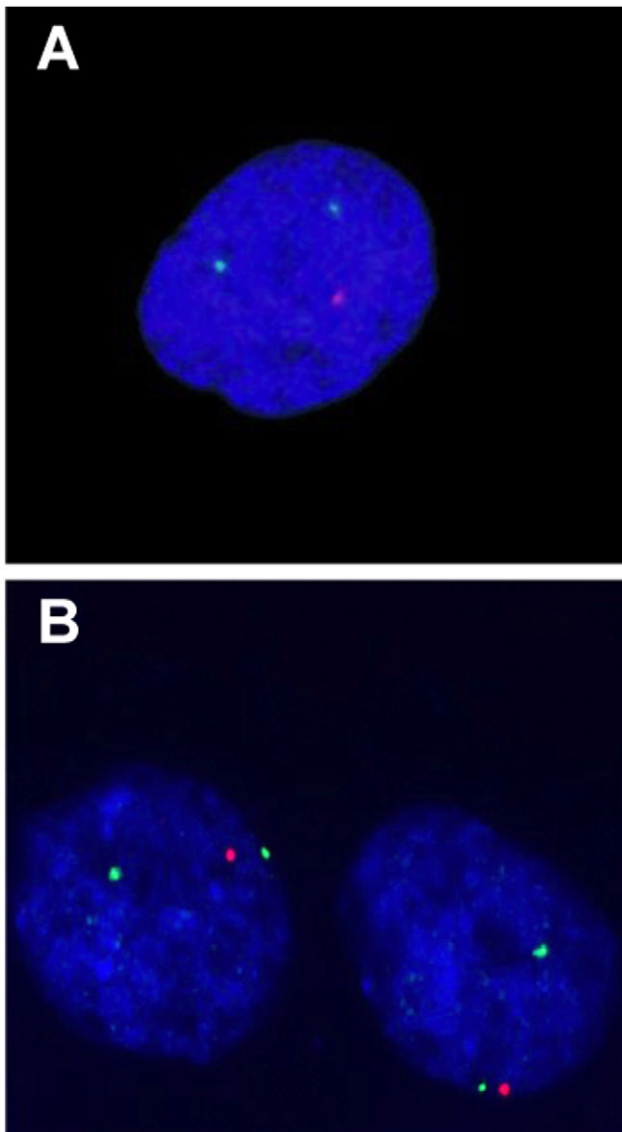
**Figure 5. Cumulative rates of gain and loss for tumors analyzed by CGH (red, 4 *BRCA2*-mutant tumors; black, 24 *BRCA2*-wildtype tumors).** A, All chromosomes; B, Chromosome 13; C, Chromosome 14. Each vertical line in B & C corresponds to an individual BAC probe. When the red line reaches  $-1$ , all of the tumors showed loss for that probe.  
doi:10.1371/journal.pone.0052079.g005

### Identification of Deletions by FISH

If the signature works by detecting large deletions on chromosomes 13 and 14, it would be better to screen tumors in a clinical setting by FISH rather than by gene expression or CGH/SNP profiling. FISH is ideally suited to detecting small changes in copy number. To test whether it would be feasible to screen for *BRCA2*-mutant tumors in this way, we performed FISH with probes mapping to the regions commonly deleted on chromosomes 13 and 14 (Figure 6). We tested nine *BRCA2* tumors and nine control *BRCA2* tumors, of which five *BRCA2* and eight *BRCA2* were not previously characterized by CGH. The results are expressed as the percentage of nuclei with less than the modal number of spots for the centromeric probes or with a ploidy of one for both probes (Table 4). The tumors were scored as “loss” when the percentage was  $\geq 50\%$ , and “other” when it was  $< 50\%$ . Contingency tables for the chromosomes individually or for both chromosomes together are shown in Table 5. For both chromosomes scored together, the sensitivity and specificity for detection of *BRCA2*-mutant tumors were 78% and 89%, respectively. We conclude that FISH provides a simple technique to screen tumors for deletions on 13q and 14q that may be associated with *BRCA2* mutations.

### Discussion

The main conclusion from our study is that deletions on chromosomes 13q and 14q are a common feature of *BRCA2*-mutant tumors. We initially set out to identify a gene expression signature that would distinguish these tumors from other tumors in patients presenting to our genetics clinics. Hierarchical clustering of the genes in the signature split the tumors into two groups in both the training and the validation sets, suggesting that the signature detects a signal that is useful for classification of the tumors. Given the GSEA and SNP/CGH results we strongly suspect that the reduced expression of the genes in the signature is caused by a reduction in the DNA copy number of the deleted regions. It is more difficult to detect deletion than amplification in gene expression data, because the former may further decrease a barely detectable signal whereas the latter can increase expression 100-fold. This probably explains why the genes in the signature are a minority of the genes in the deleted regions. Given the difficulty in measuring weakly expressed genes it is not surprising that previously reported *BRCA2* gene expression signatures did not highlight deletion of chromosomes 13 and 14 as a potential discriminating factor [19,20]. In contrast, deletion of these regions



**Figure 6. FISH with probes in the region of common deletion in a *BRCA2*-mutant tumor.** A, chromosome 13; B, chromosome 14. Red: probe in the deleted region; Green, pericentromeric probe. Each nucleus contains two green spots and one red spot, indicating that the tumor is diploid for chromosomes 13 and 14 but has heterozygous deletions in the regions tested by the red probes.  
doi:10.1371/journal.pone.0052079.g006

was noted in several previous DNA copy number and SNP studies [7,21–25]. In addition to published studies, we examined the GISTIC database (Tumorscape Release 1.6) [26] to determine whether loss of chromosomes 13 and 14 is a common event in breast cancer. Several regions are reported as harboring deletions on chromosome 13 (hg18 chr13:44680312–57088104, 57088104–114059427, 18097312–46301361 and 50901262–114059427), as expected given the presence of *BRCA2* and *RBI* on 13 q. In contrast, GISTIC reports no regions as being deleted on chromosome 14 in breast cancer at above the background rate ( $q > 0.25$ ).

There are several possible explanations for selective deletion of specific genomic regions in *BRCA2* tumors. The commonly deleted region on chromosome 13 is distal to the *BRCA2* gene, but we can not altogether exclude that *BRCA2* itself may be a driver gene in

**Table 4.** FISH with probes in the region of common deletion on chromosomes 13 and 14.

ID	BRCA status	chr 13	chr 14
52	BRCA2	84	89
86	BRCA2	90	86
106	BRCA2	100	87
133	BRCA2	93	89
A	BRCA2	84	83
B	BRCA2	100	0
C	BRCA2	87	7
D	BRCA2	100	62
E	BRCA2	100	73
16	BRCAX	0	0
F	BRCAX	0	2
G	BRCAX	0	0
H	BRCAX	100	100
I	BRCAX	4	0
J	BRCAX	0	0
K	BRCAX	2	0
L	BRCAX	7	3
M	BRCAX	10	0

Footnote. The table shows the percentage of nuclei with less than the modal ploidy or with ploidy = 1 for both the centromeric and the deletion probes. Tumours A-M were not characterized by CGH.  
doi:10.1371/journal.pone.0052079.t004

some cases, for example if there were complex genomic rearrangements on 13 q. *BRCA2* was not part of the gene signature, probably because the Affymetrix probes for *BRCA2* are not sensitive enough (the measured level was close to background and showed minimal variation). The best reporters for copy

**Table 5.** Contingency table summarizing the FISH data for deletions on chromosomes 13 and 14.

Chr 13 and 14	Other	Loss
BRCA2	2	7
BRCAX	8	1
p = 0.015		
Chr 13	Other	Loss
BRCA2	0	9
BRCAX	8	1
p = 0.0004		
Chr 14	Other	Loss
BRCA2	2	7
BRCAX	8	1
p = 0.015		

Footnote. “Loss” refers to cases where  $\geq 50\%$  of nuclei had less than the modal ploidy or had ploidy = 1. “Other” refers to cases where the value was  $< 50\%$ . The p value is for a Fisher exact test. The values for “Chr13 and 14” refer to cases where both chromosomes were affected.  
doi:10.1371/journal.pone.0052079.t005

number are housekeeping genes that lack feedback or exogenous regulation. By their nature these genes shed no light on the mechanism driving deletion. An alternative explanation is that loss of BRCA2 function generates repair intermediates or triggers checkpoint responses that are toxic in the presence of specific genes located in the deleted regions. Loss of these genes would allow the cell to resume division and form a tumor. This model predicts that the driver genes in the deleted regions should be DNA repair or checkpoint genes. *ALKBH1* could have this effect, but few other genes in the *BRCA2* signature are obvious candidates for these roles. Another possibility is that the deleted regions contain fragile sites that are more difficult to repair in the absence of BRCA2. Fragile sites are prone to replication fork collapse, a process that often leads to the formation of double strand breaks that require repair by homologous recombination. BRCA2 is required for loading of RAD51 to initiate homologous recombination [6] so increased breakage at fragile sites in the affected regions is certainly a possibility.

Screening for *BRCA2* mutations is widely performed in genetics laboratories to explain familial clustering of breast cancer. Our study design focused on patients referred to genetics clinics because this is the context in which the need to distinguish *BRCA2*-mutant from other tumors most commonly arises. Because of the size of the *BRCA2* gene it can take many months to identify mutations. This is rarely a problem in the context of genetic counseling because some interventions can be undertaken without knowledge of the mutation (for example, more frequent screening with imaging techniques) and others may even benefit from the delay by giving patients more time for reflection (for example, prophylactic mastectomy and oophorectomy). The same can not be said of medical treatment of established tumors, which must be delivered without delay. The advent of medical treatments specific for *BRCA2*-mutant tumors has created a need to identify these tumors on a more rapid time scale than has hitherto been considered necessary. In particular, *BRCA2* defects are synthetic lethal with inhibition of poly-ADP-ribose polymerase 1 (PARP1) [27,28]. We note that the *BRCA2* group in the training set contains five *BRCA2* tumors which presumably either phenocopy *BRCA2* mutation or contain *BRCA2* mutations that evaded detection by sequencing. It would be interesting to know whether tumors that phenocopy *BRCA2* mutation are also sensitive to PARP inhibitors.

In the long term it is likely that diagnostic laboratories will routinely use next generation sequencing (NGS) to identify

mutations in *BRCA2* and other relevant genes in the diagnostic biopsy when the patient initially presents with cancer. This is technically feasible but rarely performed outside major centers at present because of the cost and the complexity of the downstream bioinformatic analysis. To bridge the gap while waiting for NGS to become more widely available we propose to use FISH to screen breast tumors for deletions on 13q and 14q in order to identify tumors potentially associated with BRCA2. The technology for FISH is very well established for diagnosis of *ERBB2* amplification in sporadic breast tumors. It would require only a small modification of existing protocols to screen for loss of 13q and 14q in centers that already screen for *ERBB2* amplification by FISH. Patients whose tumors harbor deletions in those regions could then be screened by sequencing to identify either germline or somatic *BRCA2* mutations, followed by treatment with PARP inhibitors, if appropriate.

## Conclusion

We have shown that breast tumors arising in patients with germline *BRCA2* mutations have a higher frequency of deletions on 13q and 14q than is seen in other breast tumors. We propose that FISH for deletions on these chromosomes would be a rapid and technically feasible first step to enrich for tumors worth screening for *BRCA2* mutations. This would greatly facilitate the selection of patients for PARP inhibitor therapy.

## Acknowledgments

We would like to thank the patients who contributed samples to the study. We thank Véronique Fermeaux, Frédéric Bibeau and the Bergonie Cancer Institute Biological Resources Center for providing samples. We thank Marc-Henri Stern for helpful discussions; and Thierry Dubois, Marc-Henri Stern and the Translational Research Unit at the Curie Institute for providing the microarray data for the validation set. We thank the members of the Affymetrix expression array platform at IGBMC, Strasbourg.

## Author Contributions

Conceived and designed the experiments: ML RI NS. Performed the experiments: AR GB. Analyzed the data: GM RI AR EL AdR NE. Contributed reagents/materials/analysis tools: EBS ML LV IC. Wrote the paper: RI ML AR. Performed mutation screening: NJ FB NS.

## References

- Walsh T, King MC (2007) Ten genes for inherited breast cancer. *Cancer Cell* 11: 103–105.
- Fackenthal JD, Olopade OI (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. *Nat Rev Cancer* 7: 937–948.
- Collins N, McManus R, Wooster R, Mangion J, Seal S, et al. (1995) Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12–13. *Oncogene* 10: 1673–1675.
- Boulton SJ (2006) Cellular functions of the BRCA tumour-suppressor proteins. *Biochem Soc Trans* 34: 633–645.
- Greenberg RA, Sobhian B, Pathania S, Cantor SB, Nakatani Y, et al. (2006) Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. *Genes Dev* 20: 34–46.
- Yang H, Li Q, Fan J, Holloman WK, Pavletich NP (2005) The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* 433: 653–657.
- Stefansson OA, Jonasson JG, Johannsson OT, Olafsdottir K, Steinarsdottir M, et al. (2009) Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Res* 11: R47.
- Moller P, Hagen AI, Apold J, Maehle L, Clark N, et al. (2007) Genetic epidemiology of BRCA mutations—family history detects less than 50% of the mutation carriers. *Eur J Cancer* 43: 1713–1717.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98: 10869–10874.
- Banneau G, Guedj M, Macgrogan G, de Mascarel I, Velasco V, et al. (2010) Molecular apocrine differentiation is a common feature of breast cancer in patients with germline PTEN mutations. *Breast Cancer Res* 12: R63.
- R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>. Accessed 2012 Nov 22.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249–264.
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545–15550.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, et al. (2003) PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267–273.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863–14868.



18. CAPweb. Copy number Array analysis Platform on the web: A web-based platform for copy number array management and analysis, version 2.4, Available: <http://bioinfo-out.curie.fr/CAPweb/>. Accessed 2012 Nov 22.
19. Waddell N, Arnold J, Cocciardi S, da Silva L, Marsh A, et al. (2010) Subtypes of familial breast tumours revealed by expression and copy number profiling. *Breast Cancer Res Treat* 123: 661–677.
20. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bitner M, et al. (2001) Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 344: 539–548.
21. Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, et al. (2005) Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. *Cancer Res* 65: 7612–7621.
22. Jonsson G, Staaf J, Vallon-Christersson J, Ringner M, Holm K, et al. (2010) Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* 12: R42.
23. Joosse SA, Brandwijk KI, Devilee P, Wesseling J, Hogervorst FB, et al. (2012) Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. *Breast Cancer Res Treat* 132: 379–389.
24. Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, et al. (1997) Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res* 57: 1222–1227.
25. Tischkowitz M, Xia B, Sabbaghian N, Reis-Filho JS, Hamel N, et al. (2007) Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci U S A* 104: 6788–6793.
26. Beroukhi R, Getz G, Nghiemphu L, Barretina J, Hsueh T, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci U S A* 104: 20007–20012.
27. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, et al. (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434: 913–917.
28. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, et al. (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434: 917–921.