



HAL
open science

Identification de facteurs génétiques modifiant le risque de cancer chez les porteuses d'une mutation constitutionnelle d'ATM & profil tumoral des tumeurs du sein associées à une perte de fonction d'ATM.

Anne-Laure Renault

► **To cite this version:**

Anne-Laure Renault. Identification de facteurs génétiques modifiant le risque de cancer chez les porteuses d'une mutation constitutionnelle d'ATM & profil tumoral des tumeurs du sein associées à une perte de fonction d'ATM.. Biochimie, Biologie Moléculaire. Université Paris Saclay (COmUE), 2017. Français. NNT : 2017SACLS389 . tel-01721322

HAL Id: tel-01721322

<https://theses.hal.science/tel-01721322>

Submitted on 2 Mar 2018

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

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

THESE DE DOCTORAT
UNIVERSITE PARIS-SACLAY

PREPAREE DANS L'EQUIPE
EPIDEMIOLOGIE GENETIQUE DES CANCERS, INSERM U900, INSTITUT CURIE

ECOLE DOCTORALE N° 582
Cancérologie : Biologie, Médecine, Santé

SPECIALITE
Sciences de la Vie et de la Santé

Identification de facteurs génétiques modifiant le risque de cancer
chez les porteuses d'une mutation constitutionnelle d'ATM
&
Profil tumoral des tumeurs du sein associées à une perte de
fonction d'ATM.

PRESENTEE PAR
Anne-Laure Renault

SOUTENANCE
17 novembre 2017 - Institut Curie

Composition du Jury

Dr Stefan MICHIELS	Président du jury
Pr Stéphane BEZIEAU	Rapporteur
Pr Nicolas SEVENET	Rapporteur
Dr Magali OLIVIER	Examinatrice
Dr Etienne ROULEAU	Examineur
Fabienne LESUEUR	Directrice de thèse

REMERCIEMENTS

Le fruit de ces quatre années de thèse n'aurait pas été possible sans le soutien, la présence et la confiance de tout ceux qui ont participé de près ou de loin à ce projet.

Je tiens à remercier tout particulièrement Fabienne, ma directrice de thèse, pour sa confiance et sa patience, pour m'avoir guidé dans ce projet, pour tous les conseils donnés et pour m'avoir fait découvrir un peu plus le monde de la recherche. Tout thésard devrait pouvoir trouver une directrice de thèse aussi attentive et disponible !

Noura, je me demande comment auraient été les journées au labo sans toi, sûrement plus longues et moins drôles ! Merci pour ton aide, pour toutes les PCR, les confirmations de mutation, les heures passées à chercher les prélèvements sanguins, les extractions et aussi pour les fous-rires, les discussions sur le monde et pour m'avoir appris une nouvelle langue.

Je remercie toute l'équipe de m'avoir intégrée et aider au cours de ces quatre années ; Nadine pour tes conseils constructifs, Dorothée et Eve pour votre aide sur CoF-AT et pour toujours avoir pris le temps de répondre à mes questions qui nécessitaient souvent de passer par l'intranet, pour ouvrir Excel, pour ouvrir la base, pour copier dans B\$, pour copier une clé USB ! Juana, tu as eu la chance de travailler sur le superbe tableau Excel multicolore de la tumorotheque, rien que pour ça, merci ! Merci à Marie-Gabrielle et Séverine, grâce à la base TUMOPIGE que vous avez construite nous avons pu être libérées du fameux tableau Excel multicolore ! Juliette, je sais que je t'aurais beaucoup embêtée, mais je dois te l'avouer, tu chantes tellement mieux que moi, je suis un peu jalouse ...

Je n'oublie pas Elodie sans qui les analyses bioinformatiques n'auraient pas été possibles, toutes les personnes des services de génétique constitutionnelle et somatique qui nous ont fait une place dans leur locaux et qui nous ont formé sur plusieurs techniques de laboratoire.

D'un point de vue personnel, je remercie mes parents de toujours m'avoir soutenue dans mes choix, même s'ils me mènent souvent loin d'eux. Grâce à votre présence j'ai su avancer sereinement jusqu'à ce nouveau chapitre de ma vie.

Vincent, mon cœur, tu es entré dans ma vie au beau milieu de ma thèse et tu ne devais pas t'attendre à subir mes doutes, mes angoisses et mon côté ronchon. Merci de m'avoir supportée durant ces derniers mois et surtout, merci d'accepter de me suivre dans de futurs projets !

Je ne pourrais pas finir ces remerciements sans évoquer mes voisines de bureau, qui m'ont accompagnée pendant quatre ans : les toilettes. Grâce à vous je peux me reconvertir en Mme Pipi.

*« L'Homme a marché sur la lune,
mais ne sait toujours pas comment y vivre.*

*Nous avons séquencé le génome entier de l'Homme,
mais nous n'avons toujours pas percé les secrets de la vie. »*

Jean-François Mattei

Table des matières

Sommaire des figures	12
Sommaire des tableaux.....	14
Sommaire des annexes.....	15
Liste des abréviations	16
Résumé de la thèse	19
Abstract	21
<u>PARTIE 1 : INTRODUCTION.....</u>	23
1. LES CANCERS DU SEIN.....	24
1.1 RAPPEL SUR LA GLANDE MAMMAIRE ET ANATOMIE DU SEIN	24
1.2 MORPHOLOGIE DES TUMEURS DU SEIN ET FACTEURS PRONOSTIQUES	24
1.2.1 Classification histologique des adénocarcinomes	25
1.2.1.1 Carcinome canalaire invasif	27
1.2.1.2 Carcinome lobulaire invasif.....	27
1.2.1.3 Autres types d'adénocarcinomes	27
1.2.2 Facteurs histo-pronostiques	29
1.2.2.1 Bio-marqueurs anatomo-cliniques	29
1.2.2.2 Grade histologique.....	31
1.2.2.3 Récepteurs hormonaux et autres bio-marqueurs	33
1.2.3 Classification moléculaire des cancers du sein	34
1.2.3.1 Classification par expression génique.....	34
1.2.3.2 Altération du nombre de copies	34
1.2.3.3 Mutations somatiques spécifiques	35
1.2.4 Quelles caractéristiques pour quel sous-type ?.....	36
1.2.4.1 Tumeurs lumineales A.....	36
1.2.4.2 Tumeurs lumineales B.....	37
1.2.4.3 Tumeurs basales-like.....	37
1.2.4.4 Tumeurs sur-exprimant HER2	38
2. EPIDEMIOLOGIE DES CANCERS DU SEIN	39
3. LES FACTEURS DE RISQUE	42
3.1 FACTEURS ENVIRONNEMENTAUX ET CLINIQUES	42
3.1.1 Age	42
3.1.2 Antécédents personnels	42
3.1.2.1 Grossesses.....	42

3.1.2.2	Maladies bénignes du sein.....	42
3.1.3	Densité mammaire.....	43
3.1.4	Facteurs hormonaux.....	43
3.1.4.1	Facteurs hormonaux endogènes.....	43
3.1.4.2	Facteurs hormonaux exogènes.....	43
3.1.4.2.1	Contraceptifs oraux.....	43
3.1.4.2.2	Traitements hormonaux.....	44
3.1.5	Expositions aux radiations.....	44
3.1.6	Expositions liées au mode de vie.....	44
3.1.6.1	Alcool et tabac.....	44
3.1.6.2	Indice de masse corporelle.....	45
3.2	FACTEURS DE RISQUE GENETIQUES	46
3.2.1	Allèles associés à un risque élevé de cancer du sein.....	47
3.2.1.1	Gènes BRCA1 et BRCA2.....	47
3.2.1.1.1	Découverte des gènes.....	47
3.2.1.1.2	Structure des protéines BRCA1 et BRCA2.....	47
3.2.1.1.3	Fonction des protéines BRCA1 et BRCA2.....	48
3.2.1.1.4	Les allèles de prédisposition au cancer du sein.....	49
3.2.1.2	Le gène PALB2.....	51
3.2.1.3	Gènes associés à des syndromes.....	51
3.2.1.3.1	<i>TP53</i> et syndrome de Li-Fraumeni.....	51
3.2.1.3.2	<i>PTEN</i> et le syndrome de Cowden.....	52
3.2.1.3.3	<i>STK11</i> et le syndrome de Peutz-Jeghers.....	52
3.2.1.3.4	<i>CDH1</i> et les formes héréditaires de cancers gastriques.....	52
3.2.2	Allèles associés à un risque faible de cancer du sein.....	53
3.2.3	Allèles associés à un risque modéré de cancer du sein.....	54
3.2.3.1	ATM.....	55
3.2.3.2	Autres gènes de la réparation de l'ADN ayant été associés à un risque intermédiaire de cancer du sein.....	57
3.2.3.2.1	<i>CHEK2</i> et les gènes du complexe MRN.....	57
3.2.3.2.2	<i>RAD51</i>	57
3.2.3.2.3	<i>XRCC2</i> et <i>RINT1</i> , : deux nouveaux gènes de prédisposition au cancer du sein ?.....	58
3.2.3.2.4	<i>ABRAXAS</i>	58
3.2.3.3	BRIP1.....	59
3.2.3.4	Gènes impliqués dans le syndrome de Lynch.....	59

4. LA PROBLEMATIQUE ATM	61
4.1 ATM : UN GENE, UNE PROTEINE, DES FONCTIONS	61
4.1.1 La protéine ATM	61
4.1.2 Les fonctions d'ATM.....	61
4.2 L'ATAXIE-TELANGIECTASIE	62
4.3 QUELS RISQUES DE CANCER POUR LES PORTEURS HETEROZYGOTES D'UNE MUTATION D'ATM ?	63
4.4 QUELLES IMPLICATIONS EN ONCOGENETIQUE ?	63
5. OBJECTIFS DE LA THESE.....	65

PARTIE II : FACTEURS GENETIQUES ASSOCIES AU RISQUE DE CANCER DANS LES FAMILLES ATM ... 67

1. CONTEXTE DE L'ETUDE	68
2. POPULATION ETUDIEE	68
2.1 ETUDE RETROSPECTIVE SUR LES FAMILLES A-T	68
2.2 ETUDE PROSPECTIVE SUR LES FAMILLES A-T : COF-AT	69
2.2.1 La cohorte CoF-AT.....	69
2.2.2 Identification des familles A-T	69
2.2.3 L'inclusion des participantes.....	69
2.2.4 Statut mutationnel d'ATM des participantes	70
2.2.5 Suivi des participantes	70
2.3 AUTRE SOURCE D'ECHANTILLONS BIOLOGIQUES : FAMILLES A-T RECRUTEES LORS D'UNE CONSULTATION DE GENETIQUE A L'INSTITUT CURIE	71
2.4 PREVALENCE DES CANCERS DANS LES FAMILLES PARTICIPANT A COF-AT OU A RETRO-AT	71
3. RECHERCHE DE VARIANTS RARES DELETERES DANS DES GENES CONNUS DE PREDISPOSITION AU CANCER AUTRE QUE LE GENE ATM.....	73
3.1 SELECTION DES SUJETS D'ETUDE	73
3.2 PREPARATION DES BANQUES D'ADN ET SEQUENÇAGE.	77
3.2.1 Préparation des échantillons	77
3.2.2 Hybridation et amplification de la capture.....	78
3.2.2.1 Indexage et préparation au séquençage	78
3.2.2.2 Séquençage de la banque	78
3.3 ANALYSES BIOINFORMATIQUES	79
3.3.1 Traitement des données brutes.....	79
3.3.2 Identification et annotation des variants	79
3.3.3 Filtrage des variants.....	81
3.3.4 Classification des variants.....	83

3.3.4.1	Prédiction sur les substitutions faux-sens	83
3.4	RESULTATS DE L'ANALYSE SUR L'IDENTIFICATION DE NOUVEAUX VARIANTS RARES CHEZ LES INDIVIDUS SEQUENCES 84	
3.4.1	Variants « perte de fonction » identifiés	86
3.4.2	Variants faux-sens de signification clinique inconnue	91
4.	LES TELOMERES	93
4.1	STRUCTURE DE L'ADN TELOMERIQUE	93
4.2	ROLE DES PROTEINES SHELTERIN	94
4.3	REGULATION DE LA LONGUEUR DES TELOMERES	95
4.3.1	Rôle de la télomérase	95
4.3.2	Elongation des télomères	96
4.4	TELOMERES, VIEILLISSEMENT ET CANCER	96
4.4.1	Syndromes associés au vieillissement	96
4.4.2	Télomères et cancer.....	97
4.5	PROBLEMATIQUE POSEE DANS LE CADRE DE L'ETUDE SUR LES FAMILLES A-T	98
4.6	MESURE DE LA LONGUEUR DES TELOMERES CHEZ L'ENSEMBLE DES PARTICIPANTS	98
4.6.1	Choix de la technique.....	98
4.6.2	Mise en place du protocole de q-PCR.....	101
4.7	SELECTION DES ECHANTILLONS	103
5.	ROLE MODIFICATEURS DES SNPs ASSOCIES AU RISQUE DE CANCER DU SEIN ET/OU A LA LONGUEUR DES TELOMERES DANS LA SURVENUE D'UN CANCER CHEZ LES FEMMES DE LA COHORTE COF-AT OU AYANT PARTICIPEES A L'ETUDE RETRO-AT.....	104
5.1	SNPs ET LONGUEUR DES TELOMERES	104
5.2	SNPs ASSOCIES AU RISQUE DE CANCER DU SEIN OU DE L'OVAIRE	105
5.3	SNPs D'INTERET DANS LES FAMILLES A-T ET GENOTYPAGE	105
5.3.1	Sélections des SNPs d'intérêts	105
5.3.2	Génotypage.....	106
5.4	ANALYSE STATISTIQUE	108
6.	ARTICLE 1 : RESULTATS DE L'ANALYSE SUR LE ROLE DE LA LONGUEUR DES TELOMERES ET DES SNPs ASSOCIES A LA LONGUEUR DES TELOMERES SUR LE RISQUE DE CANCER CHEZ LES SUJETS HETAT.....	109
7.	ETUDE COMPLEMENTAIRE : CONTRIBUTION DE SNPs ASSOCIES AU RISQUE DE CANCER ET EN PARTICULIER AU RISQUE DE CANCER DU SEIN DANS LES FAMILLES A-T.....	120
8.	DISCUSSION	124

PARTIE III : CARACTERISATION HISTOLOGIQUE ET MOLECULAIRE DES TUMEURS DU SEIN DEVELOPPEES PAR LES PORTEURS D'UNE OU DEUX COPIES MUTEES D'ATM.125

1. CONTEXTE DE L'ETUDE	126
1.1 CARACTERISTIQUES MORPHOLOGIQUES ET MOLECULAIRES DES TUMEURS DU SEIN DEVELOPPEES PAR LES PORTEURS <i>BRCA1</i>	126
1.1.1 Caractéristiques morphologiques.....	126
1.1.2 Caractéristiques moléculaires.....	126
1.2 CARACTERISTIQUES HISTOLOGIQUES ET MOLECULAIRES DES TUMEURS DU SEIN DES PORTEURS <i>BRCA2</i>	128
1.2.1 Caractéristiques morphologiques.....	128
1.2.2 Caractéristiques moléculaires.....	128
1.3 CARACTERISTIQUES HISTOLOGIQUES ET MOLECULAIRES DES TUMEURS DU SEIN DE PORTEURS D'UNE MUTATION DANS D'AUTRES GENES DE PREDISPOSITION AU CANCER DU SEIN.	129
1.3.1 Tumeurs <i>PALB2</i>	129
1.3.2 Tumeurs <i>CHEK2</i>	130
1.3.3 Tumeurs <i>RAD51C</i>	131
1.3.4 Tumeurs <i>NBN</i>	131
1.3.5 Tumeurs <i>CDH1</i> et carcinomes lobulaires.....	131
2. PROBLEMATIQUE POSEE DANS LE CADRE DES PORTEURS DE VARIANTS D'ATM	133
3. POPULATION ET MATERIEL ETUDIES	133
3.1 FAMILLES A-T	133
3.2 FAMILLES PREDISPOSEES AU CANCER DU SEIN	134
3.2.1 Etude GENESIS	134
3.2.2 Etude kConFab	135
4. ARTICLE 2 : CARACTERISTIQUES HISTOLOGIQUES ET MOLECULAIRES DES TUMEURS ATM.	136
5. DISCUSSION	196

PARTIE IV : DISCUSSION & PERSPECTIVES.....203

1. NOUVELLES DECOUVERTES ASSOCIEES A CE PROJET DE THESE	204
2. FORCES ET LIMITES	205
3. IMPACT ET FUTUR	207

REFERENCES BIBLIOGRAPHIQUES.....211

ANNEXES.....233

Sommaire des figures

FIGURE 1 : STRUCTURE DE LA GLANDE MAMMAIRE D'APRES L'INSTITUT DU CANCER.....	24
FIGURE 2 : COUPES HISTOLOGIQUES D'ADENOCARCINOMES MAMMAIRES <i>IN SITU</i>	25
FIGURE 3 : LES DIFFERENTS SOUS-TYPE D'ADENOCARCINOMES MAMMAIRES.....	27
FIGURE 4 : ASPECT MORPHOLOGIQUE DES TUMEURS DU SEIN SELON LEUR SOUS-TYPE HISTOLOGIQUE.....	29
FIGURE 5 : TEMPS DE SURVIE DES PATIENTS EN FONCTION DU GRADE HISTOLOGIQUE.....	32
FIGURE 6 : PROFIL DES ALTÉRATIONS DU NOMBRE DE COPIES SUR L'ENSEMBLE DU GÉNOME MIS EN ÉVIDENCE PAR CGH.....	35
FIGURE 7 : INCIDENCE ET MORTALITE DU CANCER DU SEIN DANS LE MONDE ESTIMES EN 2012 PAR L'OMS.....	40
FIGURE 8 : EVOLUTION DE L'INCIDENCE ET DE LA MORTALITE PAR CANCER DU SEIN EN FRANCE METROPOLITAINE DE 1980 A 2012.....	41
FIGURE 9 : REPRESENTATION DU RISQUE RELATIF FAMILIAL DE CANCER DU SEIN EN FONCTION DU NOMBRE D'APPARENTEES ATTEINTES DE LA MALADIE.....	46
FIGURE 10 : DOMAINES FONCTIONNELS DES PORTÉINES BRCA1 ET BRCA2. A. DOMAINES FONCTIONNELS DE BRCA1.....	48
FIGURE 11 : RÉPARATION DES CASSURES DOUBLE-BRIN DE L'ADN.....	49
FIGURE 12 : MECANISMES POTENTIELLEMENT IMPLIQUES DANS LE RISQUE DE CANCER DU SEIN ET LES GENES CONNUS QUI Y SONT ASSOCIES.....	55
FIGURE 13 : PROPORTION DES CAS FAMILIAUX DE CANCER DU SEIN ET GÈNES/LOCI MIS EN ÉVIDENCE DANS LES ÉTUDES FAMILIALES ET DANS LES ÉTUDES DE POPULATION.....	60
FIGURE 14 : DOMAINES FONCTIONNELS D'ATM.....	61
FIGURE 15 : PROCESSUS DE PRÉPARATION DES ÉCHANTILLONS D'ADN AVEC LE KIT <i>SURESELECT TARGET ENRICHMENT</i> D'AGILENT.....	77
FIGURE 16 : PIPELINE BIOINFORMATIQUE UTILISÉ POUR LISTER LES VARIANTS IDENTIFIÉS DANS LES ÉCHANTILLONS SÉQUENCÉS AVEC LE PANEL DE GÈNES.....	80
FIGURE 17 : FILTRES APPLIQUÉS DANS FILEMAKERPRO POUR CLASSER LES ALTÉRATIONS IDENTIFIÉES LORS DU SÉQUENÇAGE D'UN PANEL DE GÈNES DE PRÉDISPOSITION AU CANCER.....	83
FIGURE 18 : ARBRE GÉNÉALOGIQUE SIMPLIFIÉ DE LA FAMILLE AT107.....	87
FIGURE 19 : RÉSULTATS DE L'ANALYSE DE L'INSTABILITÉ DES MICROSATELLITES DANS L'ADN TUMORAL DE LA PATIENTE AT242004.....	88
FIGURE 20 : ARBRE GÉNÉALOGIQUE SIMPLIFIÉ DE LA FAMILLE AT242.....	89
FIGURE 21 : RÉSULTATS DU SÉQUENÇAGE CIBLANT LE VARIANT C.3028G>T DE <i>MSH6</i> POUR LES PARTICIPANTES DE LA FAMILLE AT242.....	90
FIGURE 22 : ARBRE GÉNÉALOGIQUE SIMPLIFIÉ DE LA FAMILLE AT096.....	92
FIGURE 23 : RACCOURCISSEMENT DES TÉLOMÈRES AU COURS DE LA DIVISION CELLULAIRE.....	93
FIGURE 24 : REPRÉSENTATION DE LA STRUCTURE DES TÉLOMÈRES DITE EN "T-LOOP".....	94
FIGURE 25 : LE COMPLEXE DES PROTÉINES SHELTERINS.....	95
FIGURE 26 : PROTÉINES ET STRUCTURE DE LA TÉLOMÉRASE.....	96
FIGURE 27 : RÉALISATION DE LA GAMME ÉTALON.....	102
FIGURE 28 : PROFILS MUTATIONNELS DES 4 GÉNOMES À PARTIR DU CATALOGUE DE MR. STRATTON.....	198

FIGURE 29 : DIAGRAMME DE VENN DÉCRIVANT LE NOMBRE DE VARIANTS EXONIQUES SOMATIQUES COMMUNS IDENTIFIÉS PAR WGS ET PAR WES, ET LE NOMBRE DE VARIANTS IDENTIFIÉS UNIQUEMENT PAR WES EN FONCTION DU MATÉRIEL CHEZ LES PATIENTS T0077 ET T0015..... 200

Sommaire des tableaux

TABLEAU 1 : LISTE DES SOUS-TYPES HISTOLOGIQUES ET CODES DEFINIS PAR L'ORGANISATION MONDIALE DE LA SANTE.....	26
TABLEAU 2 : CLASSIFICATION DE LA TAILLE DE LA TUMEUR, DE L'ENVAHISSEMENT DES GANGLIONS TUMORAUX ET DE LA PRESENCE DE METASTASES	30
TABLEAU 3 : ÉTABLISSEMENT DU STADE TUMORAL EN FONCTION DE LA TAILLE DE LA TUMEUR, DE L'ENVAHISSEMENT GANGLIONNAIRE ET DE LA PRESENCE DE METASTASES	30
TABLEAU 4 : ATTRIBUTION DES SCORES EN FONCTION DES CARACTERISTIQUES DE LA TUMEUR	32
TABLEAU 5 : RISQUE DE CANCER DU SEIN ESTIMÉ PAR L'ANALYSE GROUPÉ CHEZ LES PORTEURS D'UN VARIANT TRONQUANT D'ATM.....	56
TABLEAU 6 : RISQUE DE CANCER DU SEIN ESTIMÉ CHEZ LES PORTEURS D'UNE SUBSTITUTION FAUX-SENS D'ATM.....	56
TABLEAU 7 : PRÉVALENCE DES CANCERS CHEZ LES PARTICIPANTS AUX ÉTUDES CoF-AT ET RETRO-AT.....	72
TABLEAU 8 : LISTE DES 55 GÈNES SÉQUENCÉS ET MALADIES HÉRÉDITAIRES ASSOCIÉES.....	74
TABLEAU 9 : LISTE DES PARTICIPANTS AYANT DÉVELOPPÉ UN CANCER CHEZ LESQUELS LES 55 GÈNES DE PRÉDISPOSITION AU CANCER ONT ÉTÉ ANALYSÉS.....	75
TABLEAU 10 : LISTE DES VARIANTS IDENTIFIÉS CHEZ LES INDIVIDUS SÉQUENCÉS AVEC LE PANEL DE 55 GÈNES.....	85
TABLEAU 11 : COMPARAISON DES DIFFÉRENTES TECHNIQUES UTILISÉES DANS LA LITTÉRATURE POUR MESURER LA LONGUEUR DES TÉLOMÈRES.....	100
TABLEAU 12 : SNPs CANDIDATS ASSOCIÉS À LA LONGUEUR DES TÉLOMÈRES DANS D'AUTRES ÉTUDES ET SÉLECTIONNÉS..	107
TABLEAU 13 : SNPs CANDIDATS ASSOCIÉS AU RISQUE DE CANCER DU SEIN DANS D'AUTRES ÉTUDES ET SÉLECTIONNÉS.....	108
TABLEAU 14 : DISTRIBUTION DES CAS DE CANCER CHEZ LES FEMMES PARTICIPANTES EN FONCTION DU STATUT MUTATIONNEL POUR ATM	120
TABLEAU 15 : ASSOCIATION ENTRE LES SNPs ASSOCIÉS AU CANCER DU SEIN ET LE RISQUE DE CANCER CHEZ LES PARTICIPANTES DE CoF-AT ET RETRO-AT.....	121
TABLEAU 16 : PUISSANCE POUR L'ÉTUDE D'ASSOCIATION ENTRE LE RISQUE DE CANCER ET LES SNPs ÉTUDIÉS.....	123
TABLEAU 17 : SIGNATURES MUTATIONNELLES IDENTIFIÉES PARMIS LES 4 GÉNOMES À PARTIR DU CATALOGUE DE SIGNATURE PUBLIÉ PAR MR. STRATTON.....	197
TABLEAU 18 : MATÉRIEL DISPONIBLE ET ANALYSES RÉALISÉES SUR LES PAIRES D'ADN NORMAL/ADN TUMORAL POUR 4 ÉCHANTILLONS.....	199
TABLEAU 19 : PROFONDEUR MOYENNE DE SÉQUENÇAGE SELON LE TYPE DE MATÉRIEL ET LE PROTOCOLE DE PRÉPARATION DES BANQUES UTILISÉES.....	200

Sommaire des annexes

Annexe 1 : <i>ABRAXAS (FAM175A)</i> and breast cancer susceptibility : no evidence of association in the breast cancer family registry.....	232
Annexe 2 : Fiche d'information pour l'étude CoF-AT (Fiche n°0).....	252
Annexe 3 : Consentement de participation à l'étude CoF-AT.....	253
Annexe 4 : Note d'information aux participantes.....	255
Annexe 5 : Premier questionnaire épidémiologique de suivi de l'étude CoF-AT.....	257
Annexe 6 : Nouvelles questions ajoutées aux questionnaires épidémiologiques de suivi numéro 1, 2, 3,4 et 5.....	288
Annexe 7 : Protocole SureSelectQXT Target Enrichment complet pour la préparation des banques d'ADN.....	293
Annexe 8 : Noms et séquences des oligonucléotides (« index ») utilisés pour identifier les échantillons pour le séquençage du panel de 55 gènes.....	301
Annexe 9 : Protocole de PCR quantitative utilisé pour la mesure de la longueur relative des télomères.....	302
Annexe 10 : Protocole de génotypage par la méthode TaqMan.....	303

Liste des abréviations

AA : Acide Aminé

ADN : Acide Désoxyribonucléique

APRAT : Association pour la Recherche sur l'Ataxie Téléangiectasie

ARN : Acide Ribonucléique

A-T : Ataxie Téléangiectasie

ATM : *Ataxia Telangiectasia Mutated*

BRCA1 : *BReast CAncer 1*

BRCA2 : *BReast CAncer 2*

CGH : *Comparative Genomic Hybridization*

CNV : *Copy Number Variation*

GAP : *Genome Alteration Print*

HetAT : porteurs hétérozygotes d'une mutation sur le gène *ATM*

HR : *Homologous Recombination*

IARC : *International Agency for Research on Cancer*

InDel : Insertion/Délétion

kb : kilo-base

LOH : *Loss Of Heterozygosity*

FAM : Fréquence de l'Allèle Mineur

Mb : mégabase

NGS : *Next-Generation Sequencing*

Non-HetAT : non-porteurs d'une mutation sur le gène *ATM*

OMS : Organisation Mondiale de la Santé

OR : *Odds Ratio*

pb : paire de bases

PIGE : Plateforme d'Investigation en Génétique et en Epidémiologie

PCR : *Polymerase Chain Reaction*

q-PCR : *quantitative Polymerase Chain Reaction*

RO : Récepteur aux Œstrogènes

RP : Récepteur à la Progestérone

RR : Risque Relatif

SNP : *Single Nucleotide Polymorphism*

SNV : *Single Nucleotide Variation*

WES : *Whole-Exome Sequencing*

WGS : *Whole-Genome Sequencing*

Résumé de la thèse

L'ataxie-télangiectasie (A-T) est une maladie génétique récessive rare de l'enfant, caractérisée par un syndrome neurodégénératif, un déficit immunitaire et des télangiectasies cutanées. La maladie est causée par des mutations bialléliques inactivatrices dans le gène *ATM* (*Ataxia-Telangiectasia Mutated*). La maladie implique aussi un risque élevé de développer des cancers, en particulier des leucémies et des lymphomes. Les sujets atteints d'A-T ont également une radiosensibilité accrue. Les femmes de plus de 50 ans apparentées à un enfant atteint d'A-T, porteuses d'une seule copie mutée d'*ATM* (HetAT), ont un risque plus élevé de cancer du sein que les femmes de la population générale (RR 4,94, 95%CI 1,90 - 12,09). Des études épidémiologiques confirment l'implication d'*ATM* dans la prédisposition au cancer du sein et montrent que 0,5% à 1% de la population porte une mutation délétère dans ce gène mais les risques de cancer pour les individus HetAT sont encore mal estimés. Dans le premier volet de ma thèse, j'ai recherché des facteurs génétiques constitutionnels pouvant modifier le risque de cancer chez les femmes de la cohorte CoF-AT (cohorte de femmes apparentées à un enfant atteint d'A-T). J'ai ensuite décrit les caractéristiques histologiques et génomiques des tumeurs du sein de sujets HetAT afin d'identifier des biomarqueurs permettant de discriminer les tumeurs ATM des autres tumeurs.

Les résultats obtenus dans la première partie de mes travaux menés sur un échantillon de 284 individus HetAT et 174 individus non-HetAT issus de 103 familles A-T montrent que les individus HetAT ont des télomères plus longs que leurs apparentés non-HetAT ($p=0.0008$). En revanche, la longueur des télomères n'est pas associée au risque de cancer dans cette population. De plus, le SNP rs9257445 (*ZNF311*) qui est associé à la longueur des télomères chez les individus HetAT n'est pas lui non plus associé au risque de cancer. En revanche les SNPs rs6060627 (*BCL2L1*) et rs2380205 (*ANKRD16*) modifient le risque de cancer chez les femmes HetAT et non-HetAT.

Les résultats obtenus dans la deuxième partie de la thèse à partir de la description morphologique de 41 tumeurs mammaires montrent que les tumeurs des porteurs d'une mutation dans *ATM* sont majoritairement de sous-type luminal B. D'un point de vue moléculaire, les 23 tumeurs ATM étudiées ne présentent pas la signature BRCAness associée à de grandes pertes chromosomiques. En revanche, nous avons montré que la majorité des tumeurs *ATM* sont tétraploïdes et présentent une perte d'hétérozygotie au locus 11q22-23 entraînant une inactivation de l'allèle normal d'*ATM* dans les tumeurs. De plus, l'analyse du nombre de copies réalisée sur ces tumeurs montre une signature *ATM* impliquant des pertes des loci 13q14.11-q14.3, 21p11.2-p11.1 et 22q11.23.

L'ensemble de ces travaux aura permis de mieux caractériser les caractéristiques génétiques des femmes de la cohorte CoF-AT et de mettre en évidence des bio-marqueurs des tumeurs *ATM*.

Abstract

Inherited biallelic mutations in the *ATM* gene cause Ataxia Telangiectasia (A-T), a multisystemic disorder characterized by neurological, cutaneous and immunological abnormalities. The disease is associated with an elevated risk of malignancies, particularly of lymphoma or leukemia, and a high radiosensitivity. Epidemiological studies have shown that female heterozygote carriers (HetAT) younger than 50 years are at increased risk of breast cancer, as compared to women from the general population (RR 4,94, 95%CI 1,90 - 12,09). Despite the rarity of A-T disease, 0.5 to 1% of the population is estimated to be HetAT. Epidemiological studies have confirmed that some specific truncating or missense variants in *ATM* are associated with increased breast cancer risk but this risk is not yet well estimated. The first part of my thesis project has consisted in characterizing inherited genetic factors modifying cancer risk in women participating in the prospective cohort CoF-AT (*“cohorte de femmes apparentées à un enfant atteint d’A-T”*). In the second part of my work, I described the morphological and molecular features of *ATM* breast tumours with the aim to identify biomarkers allowing to distinguished *ATM*-associated tumours from sporadic tumours.

Assessment of the contribution of inherited factors such as SNPs of telomere length on the risk of cancer was performed on 284 HetAT individuals and 174 non-HetAT individuals belonging to 103 A-T families. We showed that HetAT individuals have longer telomeres than their non-HetAT counterparts ($p=0.0008$). However, we found that telomere length was not associated with cancer risk in our study population. The SNP rs9257445 (*ZNF311*), which is associated with telomere length in HetAT participants, was not associated with cancer risk. Conversely, the SNPs rs6060627 (*BCL2L1*) and rs2380205 (*ANKRD16*) modified cancer risk in HetAT and non-HetAT women.

Pathology review of 41 *ATM*-associated breast tumours revealed that these tumours mostly belonged to luminal B molecular subtype. The molecular characterization of 23 *ATM*-associated tumours did not revealed the BRCAness profile associated with Large-Scale State Transitions. However, we found that *ATM* tumours were mostly tetraploid and observed loss of heterozygosity at 11q22-23 in the majority of the tumours and loss of *ATM* wild type allele. Moreover, copy number losses at loci 13q14.11-q14.3, 21p11.2-p11.1 and 22q11.23 appeared to be specific of *ATM* tumours. Altogether, this project allowed to better characterize the genetic background of the CoF-AT participants and to highlight biomarkers of *ATM* breast tumours.

Partie 1 : Introduction

1. Les cancers du sein

1.1 Rappel sur la glande mammaire et anatomie du sein

Le sein est composé de 15 à 20 lobules subdivisés en alvéoles produisant le lait, de canaux galactophores assurant la sécrétion du lait et de tissus adipeux servant de protection aux parties exocrines. A l'extérieur du sein se trouve le mamelon qui permet l'expulsion du lait et l'auréole qui sécrète de la sueur pour agir comme lubrifiant lors de l'allaitement (Figure 1).

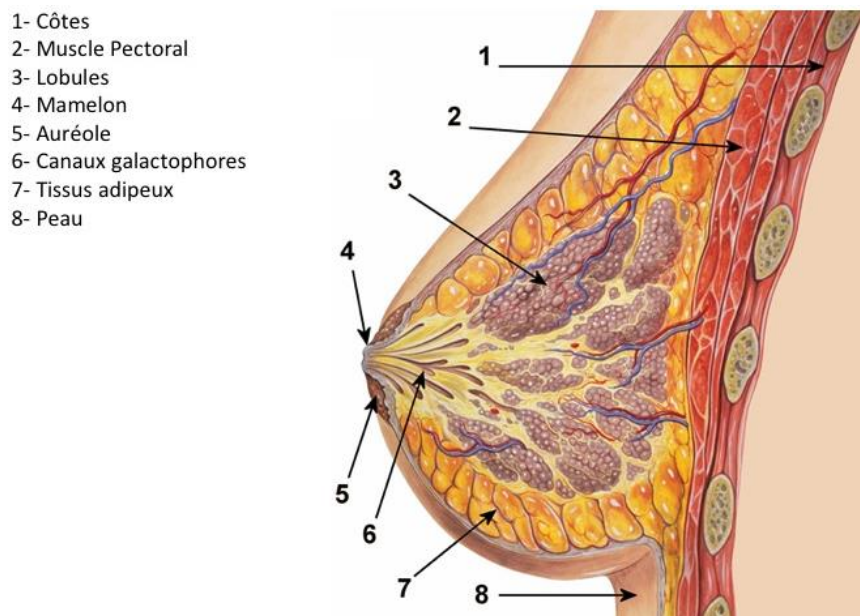


Figure 1 : Structure de la glande mammaire d'après l'Institut du Cancer.

La forme du sein évolue au cours de la vie d'une femme, car de la puberté à la ménopause les seins sont soumis aux variations des hormones sexuelles. La survenue d'un cancer entraîne également une modification de la structure du sein.

1.2 Morphologie des tumeurs du sein et facteurs pronostiques

Les tumeurs du sein, ou adénocarcinomes (du grec *adeno* signifiant glande et *carcinoma* signifiant cancer) se développent principalement à partir des cellules épithéliales. On distingue deux types principaux d'adénocarcinomes, les carcinomes *in situ* et les carcinomes invasifs, différenciables par l'état d'envahissement des cellules néoplasiques au niveau du stroma. Par ailleurs, il existe quelques rares tumeurs du sein mésoenchymateuses, fibro-épithéliales ou classées comme carcinosarcomes.

Les carcinomes *in situ* représentent 20% des carcinomes mammaires et sont définis comme une atteinte tumorale locale, de type canalaire ou lobulaire (Figure 2A et 2B), sans franchissement de la membrane basale par les cellules néoplasiques. Les carcinomes *in situ* sont sans risque de métastases et d'envahissement ganglionnaire mais sont parfois les précurseurs d'un carcinome invasif (ou « carcinome infiltrant »).

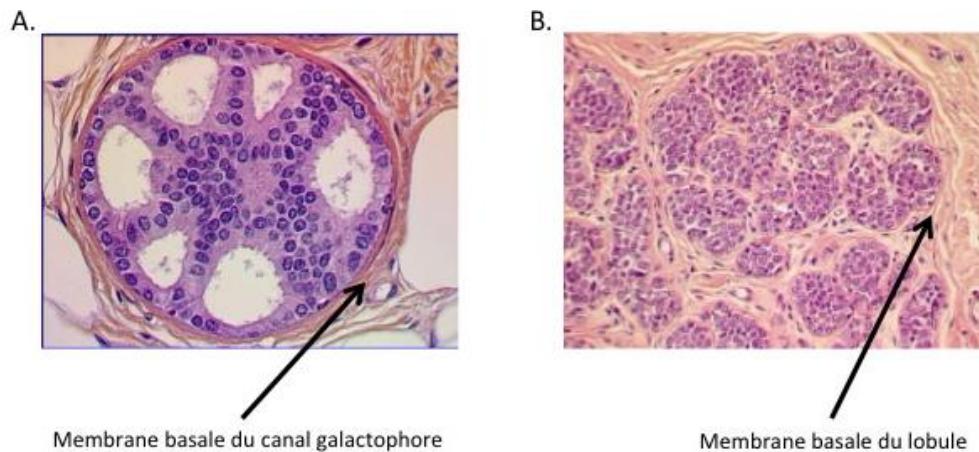


Figure 2 : Coupes histologiques d'adénocarcinomes mammaires *in situ*.

A. Adénocarcinome *in situ* canalaire. **B.** Adénocarcinome *in situ* lobulaire. Les noyaux colorés en violet correspondent aux noyaux des cellules tumorales. Les images des coupes histologiques ont été empruntées au service d'anatomopathologie de l'Institut Curie.

Les carcinomes invasifs représentent 80% des carcinomes mammaires et, à l'inverse du carcinome *in situ*, ils se caractérisent par l'envahissement des cellules néoplasiques au stroma, par un risque de métastases et d'envahissement ganglionnaire.

1.2.1 Classification histologique des adénocarcinomes

La morphologie des tumeurs du sein étant très hétérogène, l'Organisation Mondiale de la Santé (OMS) a défini des sous-types histologiques permettant de les différencier. En 2012, l'OMS a revu la classification des cancers du sein datant de 2003 et a notamment défini vingt-trois sous-types histologiques associés aux carcinomes invasifs du sein¹ (Tableau 1 et Figure 3).

Classification des carcinomes mammaires invasifs	Code de classification de la morphologie
Carcinome canalaire invasif sans atypie particulière (NST)	85003
Carcinome lobulaire	85203
Carcinome tubuleux	82113
Carcinome cribriforme	82013
Carcinome mucineux	84803
Carcinome invasif avec des caractéristiques médullaires	85103
Carcinome invasif à différenciation apocrine	84013
Carcinome invasif mésenchymateux	84903
Carcinome micro-papillaire	85073
Carcinome invasif métaplasique sans type particulier	85753
<i>Tumeurs épithélial-myoépithélial</i>	
Adénomyoépithéliome avec carcinome	89830
Carcinome adénoïde kystique invasif	82003
<i>Types rares</i>	
Carcinome invasif avec caractéristiques neuroendocrines	80133
Carcinome sécrétant	85023
Carcinome papillaire	85033
Carcinome invasif à cellules acineuses	85503
Carcinome muco-épidermoïde	84303
Carcinome polymorphe	85253
Carcinome oncocytaire	85083
Carcinome invasif à cellules riche en lipides	82903
Carcinome invasif à cellules claires, riches en glycoènes	83153
Carcinome sébacé	84103
Carcinome inflammatoire	85303

Tableau 1 : Liste des sous-types histologiques et codes définis par l'Organisation Mondiale de la Santé (d'après Sinn et al.)¹.

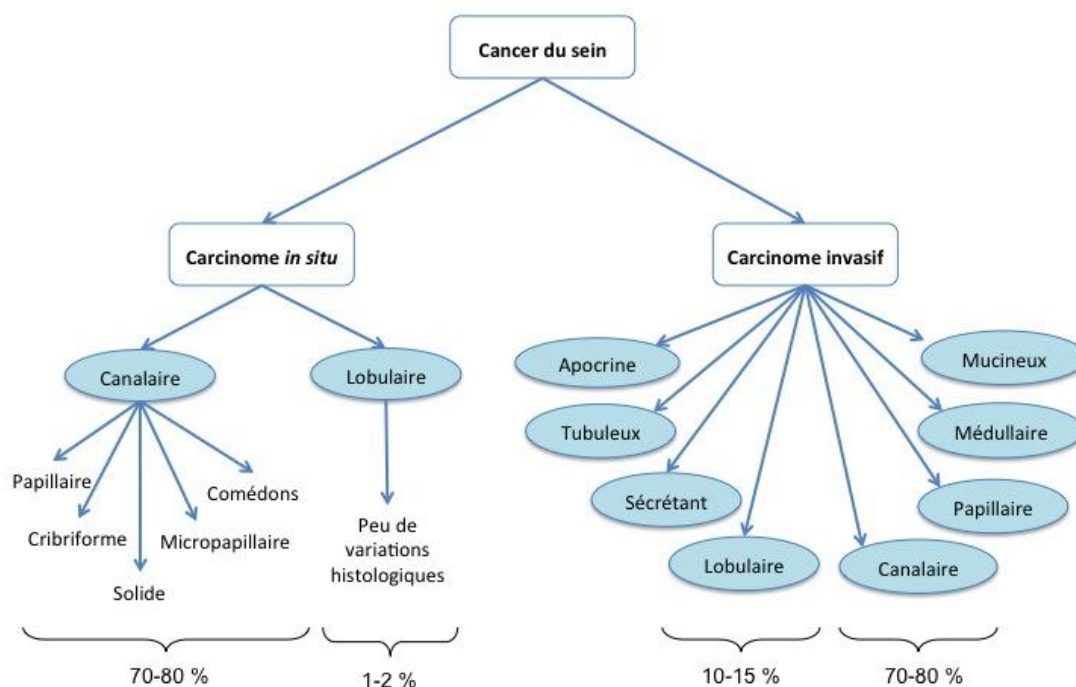


Figure 3 : Les différents sous-type d'adénocarcinomes mammaires.

Tous les sous-types histologiques rares des carcinomes invasifs ne sont pas représentés sur ce schéma mais chaque sous-type représente entre 1% et 2% des carcinomes diagnostiqués (d'après (Malhotra et al., 2010²))

1.2.1.1 Carcinome canalaire invasif

Soixante-dix pourcents des carcinomes mammaires invasifs diagnostiqués sont de type canalaire sans atypies particulières, ou « NST » (« No Special Type » en anglais) (Figure 3). D'un point de vue macroscopique, ces tumeurs forment une lésion stellaire mal délimitée (Figure 4).

1.2.1.2 Carcinome lobulaire invasif

Le carcinome lobulaire invasif est le second sous-type histologique le plus diagnostiqué, représentant 10% à 15% des carcinomes invasifs. Il se caractérise par de petites cellules néoplasiques non-adhérentes, généralement disposées en file-indienne, qui envahissent le stroma (Figure 4). Dans 90% des cas, l'absence de cohésion cellulaire dans ces tumeurs est due à l'absence d'expression de la protéine E-cadhérine³. Ce sous-type histologique est associé à un bon taux de survie à cinq ans mais cette tendance s'inverse après 6 ans de suivi des patients⁴.

1.2.1.3 Autres types d'adénocarcinomes

Différents sous-types histologiques, plus rares, se distinguent du carcinome canalaire NST par leurs atypies morphologiques. Nous allons ici en décrire quelques uns.

- Le carcinome tubuleux, représente environ 2% des tumeurs du sein et est caractérisé par une prolifération de tubules ovales ou allongés, arrangés de façon aléatoire et entourés d'une couche de tissus adipeux (Figure 4).
- Le carcinome cribiforme, identifié dans 3,5% des tumeurs du sein est reconnaissable par l'aspect criblé des cellules tumorales (Figure 4).
- Le carcinome mucineux, principalement diagnostiqué chez les femmes post-ménopausées se détache des autres sous-types histologiques par la présence d'au moins 30% de mucine extracellulaire (Figure 4).
- Le carcinome papillaire, pouvant être invasif, solide ou encapsulé, est souvent diagnostiqué chez les femmes post-ménopausées et se distingue par la présence de lésions intra-canales dessinant des projections bordées de cellules épithéliales. Il se caractérise également par une tige fribovasculaire et peu de cellules myoépithéliales^{5,6} (Figure 4).

Ces quatre sous-types histologiques sont de pronostic favorable mais d'autres sous-types rares sont associés à un mauvais pronostic, comme le carcinome apocrine ou le carcinome métaplasique. D'un point de vue macroscopique, le carcinome apocrine présente de larges cellules avec un cytoplasme granuleux et des noyaux proéminents alors que le carcinome métaplasique est un groupe hétérogène pouvant se décliner en carcinome à cellules fusiformes, en carcinome métaplasique « fibromatous-like », ou en carcinome métaplasique avec différenciation mésenchymateuse cartilagineuse, osseuse, neurogliale ou squameuse (Figure 4)⁷.

D'autre part, l'un des carcinomes mammaires le plus reconnaissable au niveau clinique est le carcinome inflammatoire, identifiable par le changement d'aspect du sein. Le sein affecté est lourd, douloureux, présentant des rougeurs, un œdème et un aspect de peau d'orange. Les tumeurs associées au carcinome inflammatoire sont souvent des carcinomes canaux NST. Il s'agit d'un sous-type histologique agressif avec un taux de survie faible des patientes^{8,9}.

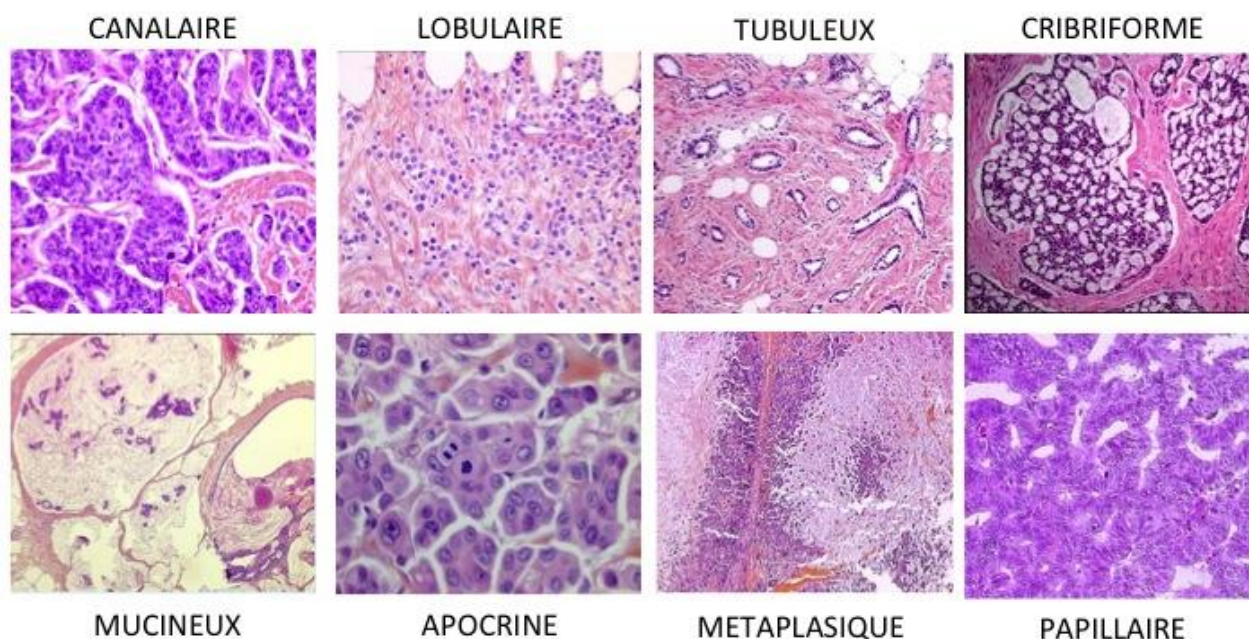


Figure 4 : Aspect morphologique des tumeurs du sein selon leur sous-type histologique.

Les différents sous-types répertoriés montrent qu'il n'existe pas une tumeur du sein mais une hétérogénéité de tumeurs (Figures 3, 4). Cette hétérogénéité tumorale est parfois présente dans une même tumeur, on parle alors d'hétérogénéité intra-tumorale.

1.2.2 Facteurs histo-pronostiques

Les sous-types histologiques donnent une indication sur le pronostic, mais il existe d'autres marqueurs morphologiques utiles pour définir le risque de rechute locale ou métastatique, ainsi que le risque de décès. Ces marqueurs permettent également de classer les carcinomes et de proposer un traitement adapté aux patients.

1.2.2.1 Bio-marqueurs anatomo-cliniques

Au niveau clinique, l'âge du patient est pris en compte dans l'estimation du pronostic : une femme de moins de 40 ans a un risque cumulé de rechute locale sur 20 ans de 34% alors que celui d'une femme de plus de 50 ans diminue à 11%¹⁰.

Au niveau anatomique, la taille de la tumeur, l'envahissement des ganglions par les cellules tumorales et la présence de métastases permettent d'évaluer le stade de propagation de la tumeur. En 1950, la classification pTNM pour « post-opératoire Taille, Nodes (ganglions en anglais), Métastases » établie par le chirurgien français Pierre Denoix est introduite par l'Union Internationale Contre le Cancer dans le système de classification des tumeurs. L'objectif était d'établir une nomenclature unique pour ces trois critères anatomiques (Tableau 2) et de décrire le stade de progression la tumeur (Tableau 3).

De façon indépendante, la taille de la tumeur permet d'évaluer les chances de survie globale à 5 ans, celles-ci étant inversement corrélées à la taille de la tumeur. Par exemple, les chances de survie des patients sont de 91% pour une tumeur de moins de 20mm et de 60% pour une tumeur de plus de 50mm. De la même manière, le taux de survie diminue en fonction du nombre de ganglions envahis par des cellules tumorales et de leur distance avec le site originel de la tumeur.

Biomarqueurs anatomiques	Classification et description du biomarqueur
Taille	pTx : non-évaluable pT0 : tumeur non-palpable pT1 : taille ≤ 2cm pT2 : 2cm ≤ taille ≤ 5cm pT3 : taille > 5cm T4 : extension de la tumeur à la paroi thoracique ou à la peau
Envahissement des ganglions régionaux	pNx : non-évaluable pN0 : absence d'envahissement pN1 : 1 à 3 ganglions envahis pN2 : 4 à 9 envahis pN3 : au moins 10 ganglions envahis
Métastases	pMx : non-évaluable pM0 : absence de métastases pM1 : présence de métastases

Tableau 2 : Classification de la taille de la tumeur, de l'envahissement des ganglions tumoraux et de la présence de métastases.

Stade de progression de la tumeur	Taille (pT)	Envahissement des ganglions régionaux (pN)	Présence de Métastases (pM)
I	pT1	pN0-pN1	pM0
	pT2	pT0	pM0
II	pT1	pN2-pN3	pM0
	pT2	pN1-pN2	pM0
	pT3	pN0-pN1	pM0
III	pT2	pN3	pM0
	pT3	pN2-pN3	pM0
	pT4	pN0-pN2	pM0
IV	pT1 à pT4	pN0 à pN3	pM1

Tableau 3 : Etablissement du stade tumoral en fonction de la taille de la tumeur, de l'envahissement ganglionnaire et de la présence de métastases.

1.2.2.2 Grade histologique

Le grade histologique, qui évalue l'agressivité de la tumeur¹¹, est l'un des éléments architecturaux de la tumeur le plus important à prendre en compte dans la prise en charge clinique du patient³. Une première définition du grade histopronostique établie par Scarff, Bloom et Richarson (classification SBR) dans les années 1960 a été révisée par Ellis et Elston (EE) en 1991¹² et cette dernière définition est utilisée comme référence. Le grade histologique est classé en trois niveaux : I - faible agressivité, III - forte agressivité et II - agressivité intermédiaire de la tumeur. Il est intéressant de noter que quelque soit la classification utilisée, SBR ou EE, la signification du grade ne change pas.

Le grade révisé par EE est déterminé par l'addition de trois critères : l'architecture de la tumeur, c'est à dire l'état de différenciation des cellules tumorales, l'anisocaryose correspondant à la forme des noyaux, et l'activité mitotique permettant d'évaluer les capacités de prolifération de la tumeur. Ces trois critères morphologiques sont scorés de 1 à 3 (Tableau 4) et la somme des scores permet de déterminer le grade de la façon suivante :

- score de 3 à 5 : grade I
- score de 6 ou 7 : grade II
- score de 8 et 9 : grade III

Le suivi de patients sur 24 mois montre que les tumeurs de bas grade sont associées à un meilleur taux de survie que les tumeurs de haut grade (Figure 5) (Histological grade, Chap 19 page 231, Ellis and Elston, Breast Pathology, S. Pinder & F. O'Malley, Elsevier Editor).

Architecture (formation de la glande mammaire)	scores
>75%	1
10% – 75%	2
<10%	3
Anisocaryose	
Léger	1
Modéré	2
Marqué	3
Index mitotiques	
0-1	1
2	2
>=3	3

Tableau 4 : Attribution des scores en fonction des caractéristiques de la tumeur

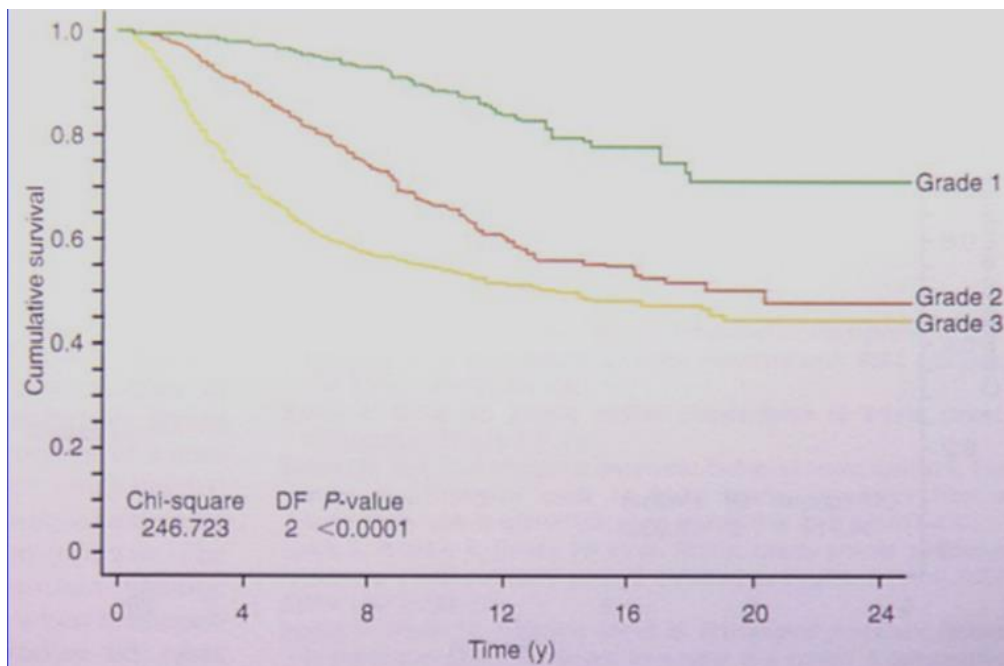


Figure 5 : Temps de survie des patients en fonction du grade histologique.

(Histological grade, Chap 19 page 231, Ellis and Elston, Breast Pathology, S. Pinder & F. O'Malley, Elsevier Editor). Les courbes de Kaplan-Meier montrent que plus le grade est élevé plus le taux de survie est faible.

1.2.2.3 Récepteurs hormonaux et autres bio-marqueurs

L'analyse de l'expression des récepteurs aux œstrogènes et à la progestérone par immunohistochimie (IHC) est utilisée comme facteur pronostic et aide à la prise en charge des patients.

Environ 60% à 80% des tumeurs expriment les récepteurs aux œstrogènes (tumeurs RO+). Les patients développant ce type de tumeurs ont de meilleures chances de survie que les patients développant des tumeurs RO⁻¹³. La présence de récepteurs aux œstrogènes à la surface des cellules tumorales oriente le traitement vers une hormonothérapie (Tamoxifène¹⁴) afin de bloquer l'action des œstrogènes sur la croissance tumorale.

Plusieurs études ont montré que l'expression des récepteurs à la progestérone était importante à prendre en compte dans l'évaluation du pronostic car les tumeurs exprimant ces récepteurs (tumeurs PR+) sont associées à une faible agressivité et à un meilleur pronostic¹⁵⁻¹⁷.

La mise en évidence en 1988 par Tal et coll. de l'amplification du proto-oncogène gène *ERBB2*¹⁸ a conduit à une meilleure prise en charge des patients. En effet, l'amplification de ce gène entraîne une surexpression de la protéine transmembranaire HER2 qui peut être ciblée spécifiquement par l'anticorps trastuzumab¹⁹. Des essais thérapeutiques ont montré une meilleure survie globale des patients bénéficiant d'un traitement à l'Herceptine (nom commercial du trastuzumab) en complément d'une chimiothérapie²⁰⁻²².

De plus, la protéine Ki-67, marqueur de prolifération cellulaire, est également utilisée comme facteur prédictif de la réponse aux traitements : plus le niveau de Ki-67 est élevé dans les cellules tumorales meilleure sera la réponse aux chimiothérapies^{23, 24}. En revanche, une forte expression de Ki-67 est associée à un faible pronostic de survie²⁵.

Nous avons vu précédemment que le pronostic diffère d'un sous-type histologique à l'autre ; cependant la prise en compte d'autres facteurs cliniques et anatomiques est essentielle dans l'évaluation du pronostic du patient.

1.2.3 Classification moléculaire des cancers du sein

Jusque dans les années 2000, seule la classification morphologique des tumeurs du sein établie par l'OMS était en vigueur. Bien qu'un suivi plus adapté était proposé aux patients développant une tumeur triple-négative, peu de données permettaient de classer plus finement les tumeurs afin d'adapter les traitements aux autres types de tumeurs.

Grâce aux avancées technologiques, une meilleure compréhension de la taxonomie des cancers du sein a pu être établie permettant *in fine* de tendre un peu plus vers une médecine « de précision ». Différents groupes de tumeurs se ressemblant ont été mis en évidence selon leurs caractéristiques moléculaires, ce qui a permis de définir des protocoles thérapeutiques ciblant par exemple certains types d'altérations génomiques.

1.2.3.1 Classification par expression génique

En 2000, Perou et coll. émettent l'hypothèse que la diversité phénotypique des tumeurs est liée à une diversité d'expression génique²⁶. Pour investiguer cette hypothèse ils mesurent dans 40 tumeurs du sein (36 carcinomes canauxiaux invasifs, deux carcinomes lobulaires invasifs, un carcinome canalaire *in situ* et un fibroadénome) l'expression de plus de 8 000 gènes au moyen d'une puce ciblant l'ADN complémentaire (cDNA) de ces gènes. Pour la première fois dans la littérature, les termes de « portrait moléculaire » et de « sous-type intrinsèque » émergent. Quatre sous-types intrinsèques se distinguent : les tumeurs « ER+/luminales », sur-exprimant les gènes normalement exprimés par les cellules luminales ; les tumeurs « basales-like », associées à une forte expression des gènes exprimés dans les cellules basales ; les tumeurs « ERBB2+ » sur-exprimant ce gène, et les tumeurs « normal-like » ayant un profil d'expression similaire à celui du tissu mammaire sain. Un an plus tard, la même équipe confirme ces résultats dans une deuxième série de 78 tumeurs mammaires (71 carcinomes canauxiaux invasifs, cinq carcinomes lobulaires invasifs, et deux carcinomes canauxiaux *in situ*) et dissocie les tumeurs ER+/luminales en classe A et B²⁷.

Par ailleurs, une autre étude, portant sur des sous-types histologiques plus rares montre qu'il est possible d'associer un sous-type histologique à un sous-type moléculaire²⁸. Par exemple, les carcinomes médullaires ont un profil d'expression « basal-like » alors que les carcinomes tubuleux appartiennent au sous-type luminal.

1.2.3.2 Altération du nombre de copies

Il est aujourd'hui bien établi que les tumeurs se développent à partir de l'accumulation d'aberrations génétiques ou épigénétiques et que des altérations spécifiques récurrentes permettent de distinguer des groupes de tumeurs particuliers. Dès 1992, Kallioniemi et coll. montraient par hybridation

fluorescente *in situ* (FISH) dans des cellules tumorales du sein que même en cas de perte d'hétérozygotie (« LOH » pour Loss Of Heterozygosity)²⁹ au niveau du locus 13q14 contenant le gène *RB1*, deux copies de ce gène étaient pourtant présentes dans les cellules tumorales. Hicks et *al.* ont ensuite montré avec la technique d'hybridation génomique comparative (« CGH » pour *Comparative Genomic Hybridation*) que les variations du nombre de copies (« CNV » pour *copy number variation*), identifiées sur l'ensemble du génome permettaient de classer les tumeurs en 3 groupes : un profil d'altérations simple (ou « simplex ») caractérisé par peu de régions d'amplifications et de pertes ; un profil « complexe I » ayant une apparence en dent de scie et un profil « complexe II » présentant une alternance d'amplifications et de délétions (Figure 6)³⁰. Plusieurs études, dont l'étude METABRIC et l'étude du Cancer Genome Atlas (TCGA), réalisées sur de plus larges séries de tumeurs, ont confirmé la présence de profils CNV spécifiques aux différents sous-types moléculaires de tumeurs du sein³¹⁻³³.

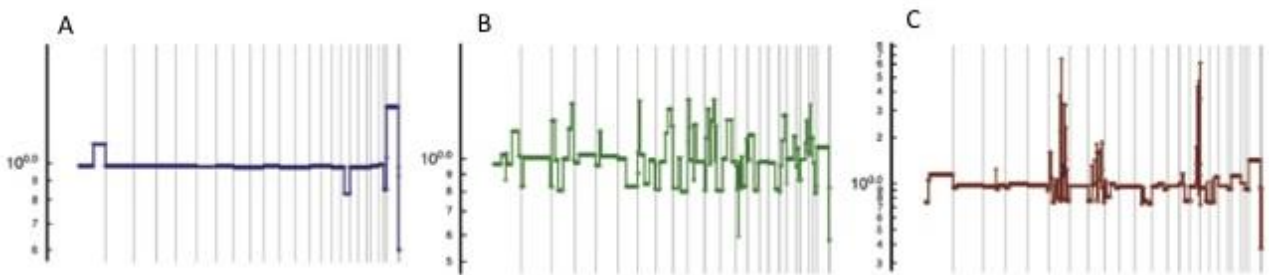


Figure 6 : Profil des altérations du nombre de copies sur l'ensemble du génome mis en évidence par CGH.
A. Profil « simplex » B. Profil « complexe I » C. Profil « complexe II » d'après Hicks et coll³⁰.

1.2.3.3 Mutations somatiques spécifiques

L'essor du séquençage à très haut débit a affiné les connaissances sur la nature des mutations somatiques (substitutions, insertions, délétions, réarrangements) pouvant conduire à l'évolution de la tumeur, sur la localisation de ces mutations dans le génome et sur leur fréquence. Dans les tumeurs du sein, une dizaine de gènes se révèlent fréquemment mutés : *TP53* (*Tumor Protein 53*), *PIK3CA* (*PIK3CA phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha*), *MYC* (*MYC proto-oncogene*), *CCND1* (*Cyclin D1*), *PTEN* (*Phosphatase and tensin homolog*), *FGFR1* (*Fibroblast growth factor receptor 1*), *GATA3* (*GATA binding protein 3*), *RB1* (*RB transcriptional corepressor 1*), *ERBB2* (*erb-b2 receptor tyrosine kinase 2*) et *MAP3K1* (*mitogen-activated protein kinase kinase kinase 1*)^{31, 33-35}. Ces gènes mutés caractéristiques des tumeurs du sein ne sont cependant pas tous des cibles thérapeutiques. De plus, le séquençage de nombreux génomes tumoraux mammaires a permis de mettre en évidence de nouvelles cibles thérapeutiques visant les voies biologiques impliquant les

gènes *BRAF* (*BRAF proto-oncogène*), *mTOR* (*mechanistic target of rapamycin kinase*) ou encore *EGFR* (epidermal growth factor receptor)³⁶.

Par ailleurs, le séquençage du génome de 560 tumeurs du sein par l'équipe de MR. Stratton au Wellcome Trust Sanger Institute en Angleterre a permis d'identifier 12 « signatures » reposant sur une prépondérance de transversions (C>T ; C>A) ou de transitions (A>G ; T>C). Des mutations localisées dans 93 gènes du cancer et potentiellement impliquées dans le développement tumoral (mutations « *driver* » ou « conductrices ») ont également été mises en évidence³⁴.

Ainsi, de nombreuses données sur les mutations somatiques ont été générées durant cette dernière décennie. L'enjeu actuel est de déterminer quelles mutations sont de potentielles mutations « *driver* » dans le développement du cancer pour mieux comprendre leur rôle dans l'étiologie de la maladie et ainsi identifier des cibles thérapeutiques potentielles.

Les catalogues de mutations issus des projets tels que The Cancer Genome Atlas (TCGA), le International Cancer Genome Consortium (ICGC) ou bien les signatures établies par le Wellcome Trust Sanger Institute sont mis à disposition de la communauté scientifique pour répondre à des questions particulières. En effet, comme nous le verrons dans le chapitre portant sur la caractérisation moléculaire des tumeurs du sein des porteurs d'une mutation *ATM*, ces données permettent, par exemple, de comparer deux séries de tumeurs pour déterminer si certaines mutations sont surreprésentées dans un groupe particulier de tumeurs. Je n'évoque ici que les cancers du sein, mais de telles données sont également disponibles pour de nombreux autres types de cancer tel que les cancers de la prostate, du pancréas ou encore le mélanome.

1.2.4 Quelles caractéristiques pour quel sous-type ?

Les sous-types moléculaires définis par Perou et *coll.* ont pu être précisés par l'analyse du transcriptome, la CGH et le séquençage à haut débit. Certaines caractéristiques moléculaires sont aujourd'hui clairement identifiées, bien qu'il ne s'agisse encore que d'une liste non-exhaustive d'altérations.

1.2.4.1 Tumeurs lumineales A

Environ 40% des carcinomes canauxiers invasifs sont de sous-type luminal A³⁷. Ces tumeurs sont RO+, généralement de bas grade avec une faible prolifération cellulaire et sont de pronostic favorable²⁷. Ce sous-type moléculaire est caractérisé par une surexpression du gène *ESR1* (codant le récepteur aux œstrogènes) ainsi que des gènes régulés par cette hormone, tels que *GATA3*, *SLC39A6* (*Solute carrier family 39 member 6*), *XBP1* (*X-box binding protein*) ou *FOXA1* (*Forkhead box A1*)³⁸.

L'analyse des variations du nombre de copies (CNVs) montre que les tumeurs lumineales A présentent un profil peu remanié (Figure 6.A). Cependant, l'amplification des bras chromosomiques 1q et 16p et la perte du bras 16q sont des événements conjoints fréquemment trouvés dans ces tumeurs³².

Peu de mutations somatiques de type SNV (« single nucleotide variations » ou variations nucléotidiques simples) ont été identifiées dans les tumeurs lumineales A. Néanmoins les gènes *PIK3CA*, *GATA3*, *MAP3K1*, *KMT2C* (*lysine methyltransferase 2C*) et *CBFB* (*core-binding factor beta subunit*) s'avèrent souvent altérés dans ces tumeurs.

1.2.4.2 Tumeurs lumineales B

Les tumeurs lumineales B représentent 20% des carcinomes canalairens invasifs et appartiennent au groupe des tumeurs RO+. Comme les tumeurs lumineales A, les tumeurs lumineales B présentent une expression élevée des gènes régulés par l'œstrogène mais sont en revanche de moins bon pronostic²⁷. De plus, elles se distinguent par une forte expression des gènes de prolifération cellulaire tels que *v-MYB* (*MYB proto-oncogene*), *GGH* (*Gamma-glutamyl hydrolase*), *LAPTMB4* (*Lysosome-associated transmembrane protein 4 beta*), *NSEP1* (*Nuclease sensitive element binding 1*), and *CCNE1* (*Cyclin E1*)³⁹, ainsi que par la récurrence des mutations somatiques dans les gènes *PIK3CA* et *TP53*⁴⁰.

Les profils CNVs des tumeurs lumineales B sont de type « complexe II » (Figure 6.C), avec une alternance de gains et de pertes. Des amplifications des loci 1q et 8q associées à des pertes des loci 1p, 8p, 16q, 17q et 22q sont souvent mises en évidence dans ces tumeurs^{33, 41, 42}.

1.2.4.3 Tumeurs basales-like

Ce sous-type moléculaire représente environ 15% des carcinomes canalairens invasifs et est associé aux tumeurs RO-, plus particulièrement aux tumeurs dites triple-négatives n'exprimant pas les récepteurs hormonaux ni la protéine HER2 (RO-, PR-, HER2-). Les tumeurs basales-like sont de haut grade et sont associées à un mauvais pronostic, ce qui est probablement en lien avec la prolifération accélérée des cellules tumorales⁴³. Bien que ces tumeurs ont un profil mutationnel très hétérogène, on note cependant un fort taux de mutations de *TP53*^{31, 44}. Le profil CNV de ce sous-type moléculaire est très caractéristique avec de nombreuses amplifications et délétions (profil « complexe I » en dent de scie)³ (Figure 6.B).

Par ailleurs, les tumeurs basales-like présentent une déficience de la réparation des dommages à l'ADN par recombinaison homologue. La découverte de cette vulnérabilité des cellules a permis l'émergence de deux classes de nouvelles molécules, les cisplatines et les inhibiteurs de la poly ADP-ribose polymérase 1 (PARP)⁴⁵ (Olaparib), majorant l'anomalie de réparation des cassures double-brin de l'ADN par recombinaison homologue et faisant l'objet de nombreux essais cliniques.

1.2.4.4 Tumeurs sur-exprimant HER2

Les tumeurs HER2+ sont souvent de haut grade et agressives, d'où le mauvais pronostic qui leur est associé⁴⁶. Parmi les tumeurs HER2+, on distingue deux sous-groupes de tumeurs en fonction de l'expression, ou non, des récepteurs aux œstrogènes.

Les tumeurs RO+/HER2+, également appelées lumorales B/HER2+, présentent les mêmes caractéristiques que les tumeurs lumorales B au niveau de l'expression de *ESR1* et des gènes régulés par l'œstrogène tandis que les tumeurs RO-/HER2+ se distinguent des tumeurs lumorales B/HER2 par l'absence d'expression des gènes associés au sous-type luminal et par un taux plus élevé de mutation de *TP53*³.

Les tumeurs HER2+ forment un groupe très hétérogène, aucun profil génomique ou mutationnel clair n'ayant été associé à ce sous-type moléculaire.

En 2015, des experts réunis à Vienne à la quatorzième conférence internationale de St Gallen sur le cancer du sein ont proposé une nouvelle méthode de classification des tumeurs permettant de faire une approximation des sous-types intrinsèques moléculaires à partir des résultats d'immunohistochimies (IHC) obtenus en routine par les laboratoires d'anatomopathologie^{47, 48} :

- Tumeurs lumorales A : RO+, RP+/-, HER2- et Ki-67<20%
- Tumeurs lumorales B : RO+, RP+/-, HER2- et Ki-67≥20%
- Tumeurs lumorales B/HER2+ : RO+, RP+/-, HER2+ et Ki-67≥20%
- Tumeurs triple-négatives : RO-, RP-, HER2-
- Tumeurs HER2-positives : RO-, RP-, HER2+

A partir des données d'expression, de l'analyse des CNVs et de l'identification de mutations somatiques récurrentes, des outils de diagnostic tel que les tests PAM50 (Prediction Analysis of Microarrays)^{49, 50}, MammaPrint⁵¹, Oncotype DX assay⁵², Breast cancer Index⁵³ ou encore EndoPredict⁵⁴ sont aujourd'hui utilisés pour prédire la réponse aux traitements, le risque de rechute ou les bénéfices d'une hormonothérapie étendue à 5 ans pour ainsi améliorer la prise en charge des patients.

La classification des cancers du sein est devenue de plus en plus intégrative avec l'accumulation des connaissances au niveau moléculaire. Le diagnostic et la prise en charge des patients repose donc sur une approche multidisciplinaire, mettant en relation les oncologues, les anatomopathologistes, les plateformes de génomique, les laboratoires de diagnostic moléculaire et les généticiens.

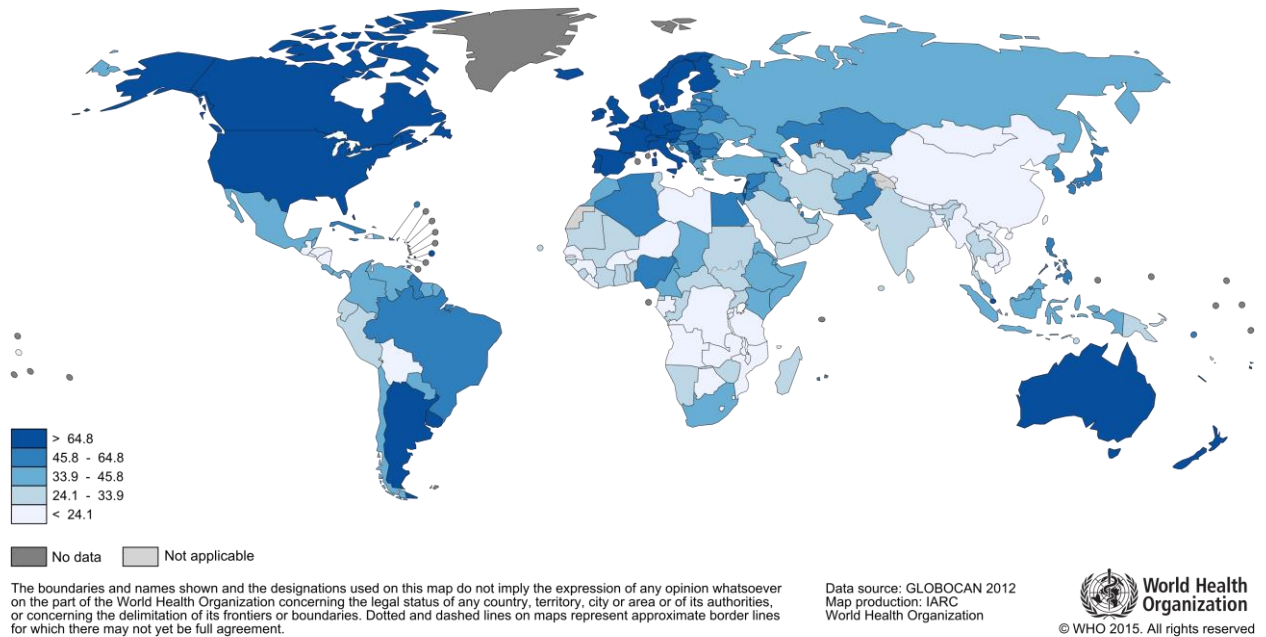
2. Epidémiologie des cancers du sein

Selon les données de l'*International Agency for Research on Cancer* (IARC) (<http://globocan.iarc.fr>) le cancer du sein est le cancer le plus fréquent chez les femmes dans le monde avec 1,67 millions de nouveaux cas diagnostiqués en 2012⁵⁵. Bien que ce type de cancer affecte majoritairement les femmes, environ 1% des cas sont des hommes. Le cancer du sein est la cinquième cause de mortalité par cancer, touchant aussi bien les pays développés que les pays dits « en voie de développement ». Cependant, une incidence et un taux de survie plus élevés sont observés dans les pays développés comme l'indiquent les données d'Amérique du Nord (incidence : 92 pour 100 000 personnes, taux de survie : 60%) et d'Afrique (incidence 24 pour 100 000 personnes, taux de survie : 40%) (Figure 7)⁵⁵.

L'étude de l'IARC de 2013, portant sur les données de registre de 40 pays estimait l'incidence du cancer du sein en France en 2012 à 137/100 000 femmes et un taux de mortalité de 24/100 000 femmes⁵⁶. En 2015, d'après l'Institut National du Cancer (INCa), 54 000 nouveaux cas ont été diagnostiqués et la maladie a été la première source de mortalité par cancer avec 11 900 décès recensés (données de « La situation du cancer en France en 2015" l'Institut National du Cancer). Néanmoins entre 2005 et 2012, une diminution de 1.5% par an de l'incidence et de 1.5% par an de la mortalité ont été observées (Figure 8)⁵⁷. La diminution de l'incidence et de la mortalité peut s'expliquer par la généralisation du dépistage organisé en France depuis 2004. Ce dépistage organisé consiste en une mammographie réalisée tous les deux ans pour les femmes âgées de 50 à 74 ans ne présentant pas de prédispositions génétiques connues. La proportion de femmes diagnostiquées grâce au dépistage est estimée, selon l'INCa, à 10-15% des cas diagnostiqués par an.

La survie nette standardisée sur l'âge s'améliore également depuis quelques années. En effet, selon les données de registre du réseau FRANCIM, la survie à 5 ans est passée de 80% entre 1989 et 1993 à 87% entre 2005 et 2012 (http://www.e-cancer.fr/ressources/cancers_en_france/). La survie nette à 10 ans est de 76% (données de l'INCa, 2010). Cette évolution s'explique par l'amélioration des thérapeutiques et de la prise en charge des patients.

A.



B.

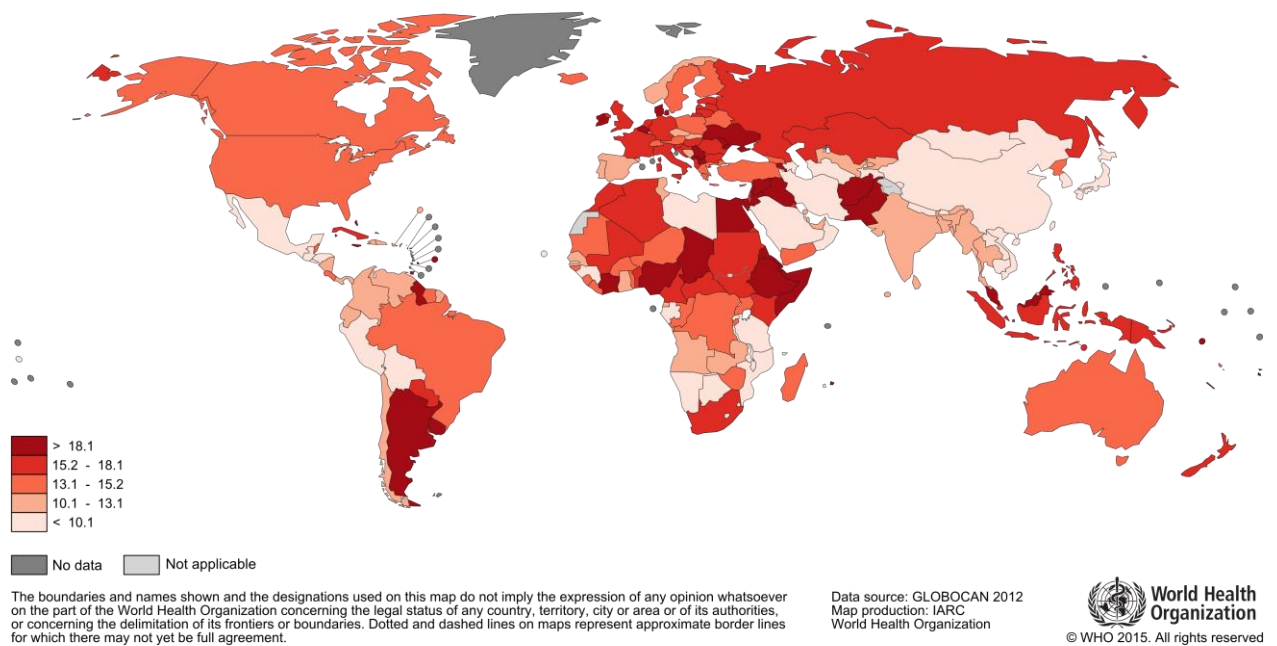


Figure 7 : Incidence et mortalité du cancer du sein dans le monde estimés en 2012 par l'OMS.

A. Incidence du cancer du sein. B. Mortalité liée au cancer du sein.

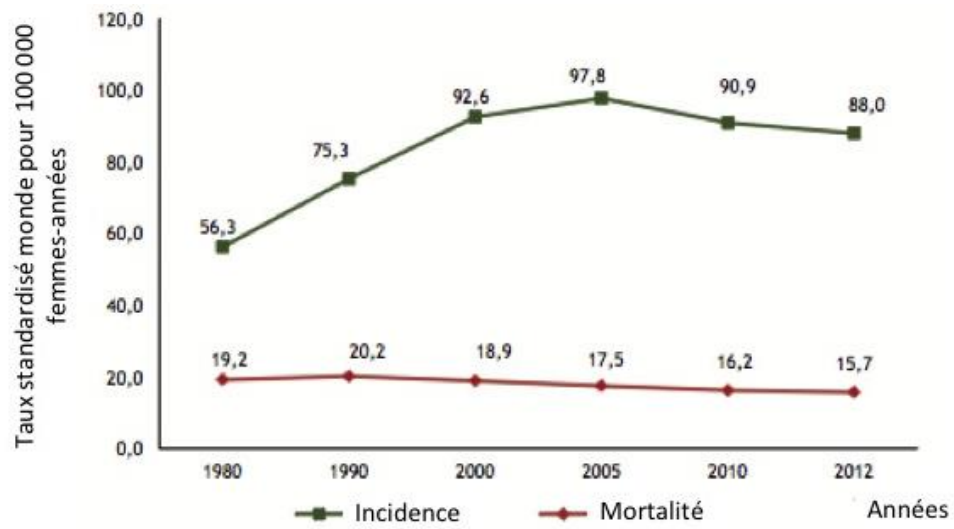


Figure 8 : Evolution de l'incidence et de la mortalité par cancer du sein en France métropolitaine de 1980 à 2012. (d'après Binder-Foucard, 2013)⁵⁷.

3. Les facteurs de risque

Au cours de sa vie une femme est soumise à différentes expositions environnementales ou biologiques qui peuvent influencer sur le risque de développer un cancer du sein.

3.1 Facteurs environnementaux et cliniques

3.1.1 Age

L'âge est l'un des facteurs de risque les plus importants. En effet, le risque de cancer du sein pour une femme de la population générale augmente par tranche de 10 ans : ce risque est de 0,44% pour une femme de 30 ans, sans histoire familiale de cancer du sein et de 3,82% à 70 ans (Howlader N, Noone AM, Krapcho M, et. al. (eds). SEER Cancer Statistics Review, 1975–2009 (Vintage 2009 Populations)).

3.1.2 Antécédents personnels

Différents aspects de la vie d'une femme influencent le risque de développer un cancer du sein, tel que l'âge à la première grossesse ou encore les antécédents médicaux.

3.1.2.1 Grossesses

De façon générale, une femme ayant mené une grossesse à terme avant ses 30 ans présente un risque de cancer du sein moins élevé qu'une femme nullipare⁵⁸. Le nombre de grossesses menées à terme est également un facteur protecteur du risque de développer la maladie, celui-ci diminuant de 3% à chaque grossesse⁵⁹. Une étude prospective norvégienne réalisée chez plus d'un million de participantes a montré que l'effet protecteur de la parité était plus fort chez les femmes ayant eu une première grossesse avant 20 ans comparé aux femmes ayant accouché d'un premier enfant après leur 30 ans⁶⁰. Bien que l'âge à la première grossesse menée à terme et le nombre de grossesse soient des facteurs protecteurs du risque de cancer du sein, une augmentation transitoire de ce risque après la première grossesse a été mise en évidence⁶¹. L'effet protecteur de la parité n'apparaît qu'à l'âge de 40 ans pour les femmes ayant mené une première grossesse à terme avant 25 ans, et à environ 50 ans pour les femmes ayant mené une première grossesse à terme entre 25 et 29 ans⁶¹.

3.1.2.2 Maladies bénignes du sein

Les maladies bénignes du sein telles que les fibroadénomes, les kystes, les écoulements au niveau du mamelon ou encore les hyperplasies augmentent le risque de cancer du sein. Une étude de suivi portant sur 9 087 femmes a montré que les femmes développant une hyperplasie atypique du sein

avaient un risque de cancer du sein environ quatre fois supérieur à celui de femmes non atteintes d'une maladie bénigne du sein (RR=4,24 (95% IC : 3,26 - 5,41))⁶².

3.1.3 Densité mammaire

A l'imagerie, la densité mammaire absolue est définie comme le rapport entre le tissu glandulaire et le tissu adipeux. Les seins sont denses lorsque le tissu glandulaire est la composante majoritaire du sein. En 2006, McCormack et *al.* montrent à partir des données de plus 14 000 cas et 226 000 témoins que le risque de cancer du sein augmente avec la densité mammaire⁶³. Les femmes ayant entre 25% et 49% de tissus glandulaires ont un risque deux fois plus élevé que les femmes ayant moins de 5% de tissus glandulaires (RR 2.1 (95% IC : 1.70, 2.63)); ce risque est multiplié par cinq pour les femmes ayant plus de 75% de tissu glandulaire (RR 4.6 (95% IC : 3.64- 5,91)). Récemment il a été montré que le pourcentage de zones denses est plus important à prendre en compte dans l'estimation du risque que la densité absolue⁶⁴.

3.1.4 Facteurs hormonaux

Les facteurs hormonaux décrits comme ayant une influence sur le développement des tumeurs du sein sont de deux types : les facteurs endogènes, qui concernent les hormones produites par notre système endocrinien, et exogènes, comme les hormones contraceptives ou les traitements hormonaux de substitution (THS) utilisés lors de la ménopause.

3.1.4.1 Facteurs hormonaux endogènes

La glande mammaire est soumise au cours des cycles menstruels et pendant les grossesses aux variations hormonales. Dès 1932, Antoine Lacassagne rapporte dans le compte rendu de l'Académie des Sciences, que la folliculine, un type d'œstrogène, est impliqué dans le développement de tumeur mammaire chez la souris⁶⁵. Depuis, plusieurs études ont montré qu'une exposition plus longue aux hormones sexuelles liée à l'apparition des règles avant l'âge de 12 ans ou à une ménopause tardive augmente le risque de cancer du sein^{58,66}.

3.1.4.2 Facteurs hormonaux exogènes

3.1.4.2.1 Contraceptifs oraux

La mise en évidence d'un lien entre la prise de contraceptifs oraux et le risque de cancer du sein est un réel enjeu de société car 10% des femmes en âge de procréer (environ 100 millions) utilisent un moyen contraceptif. La méta-analyse du *Collaborative Group on Hormonal Factors in Breast Cancer* réalisée en 1996 chez 61 196 femmes utilisatrices d'un contraceptif oral et 90 747 non-utilisatrices,

atteintes ou non de cancer du sein, montrent une faible augmentation du risque de cancer du sein chez les utilisatrices d'un contraceptif oral (RR=1,07, 95%IC : 1,02 - 1,13)⁶⁷. Ce risque décroît à l'arrêt de la contraception, jusqu'à atteindre le risque associé à l'âge de la femme⁶⁷. Par la suite, les données combinées de la cohorte du *Women's Health Life* et de deux études de suivi norvégiennes et suédoises regroupant plus 103 000 femmes ont permis d'estimer le risque de cancer du sein à partir de 1 800 cas prospectifs. Selon cette étude, le risque dû à la prise de contraceptifs serait augmenté de 1,6 (95% IC : 1,0 – 2,4) mais il n'y aurait pas de variation de ce risque en fonction de la durée d'exposition⁶⁸. En 2002 et 2003 deux autres études de suivi menées dans la population américaine ne montrent pas d'association entre la prise de contraceptif oral et le risque de cancer du sein^{69, 70}. Malgré des résultats divergents, l'IARC a classé en 2015 les contraceptifs oraux dans le « groupe 1 » des agents carcinogènes, ce qui les place au rang des agents carcinogènes pour l'Homme⁷¹.

3.1.4.2.2 Traitements hormonaux

Les THS sont prescrits aux femmes lors de la ménopause pour diminuer progressivement la dose d'hormones dans le sang afin de contrer les symptômes de la ménopause, comme les modifications osseuses, les douleurs articulaires ou musculaires, les bouffées de chaleur ou des manifestations neuropsychiques liées à la privation œstrogénique. De façon générale, les femmes exposées aux THS ont un risque augmenté de cancer du sein de 26% à 35% comparé aux femmes non-exposées.

L'un des traitement hormonal concentrant le plus de preuves d'association avec le risque de cancer du sein, est le diéthylstilbestrol ou Distilbène (DES), prescrits aux femmes de 1950 à 1970 en prévention des fausses couches et des risques de prématurité. La prise de ce médication est associée à une augmentation du risque de 1.4 (95% IC : 1,1 - 1,9) chez les femmes ayant pris ce traitement et de 1,91 (95% IC : 1,09 - 3,33) chez leurs filles^{72, 73}.

3.1.5 Expositions aux radiations

Les radiations ionisantes (rayons X, rayons gamma) ont été associées à une augmentation du risque de cancer du sein dans des études s'intéressant à l'impact des ces radiations^{74, 75}. De plus, trois études portant respectivement sur 105⁷⁶, 107⁷⁷ et 16⁷⁸ cas ont évalué l'impact d'une radiothérapie sur le risque de développer un cancer du sein en deuxième localisation et ont montré un effet linéaire entre la dose reçue et l'augmentation du risque.

3.1.6 Expositions liées au mode de vie

3.1.6.1 Alcool et tabac

La consommation d'alcool, de tabac ou encore l'indice de masse corporel (IMC) sont des facteurs influençant le risque de cancer du sein. L'alcool et le tabac sont classés comme agents carcinogènes

pour l'Homme (groupe 1 des agents carcinogènes par l'IARC). Une consommation journalière de plus de 3 à 4 verres augmente le risque de cancer du sein de 1,32 (95% IC : 1,19 - 1,45) par rapport aux femmes qui déclarent ne pas consommer d'alcool, ce risque augmente à 1,46 (95% IC : 1,33 - 1,61) pour les femmes consommant plus de 4 verres/jour⁷⁹.

Les expositions au tabac, quelles soient actives ou passives ont un impact négatif sur le risque de cancer du sein comme sur bien d'autres cancers. Une étude de suivi réalisée sur plus de 50 000 sœurs a montré que les fumeuses avaient un risque de cancer du sein augmenté de 17% (95% IC : 1,00 - 1,36) par rapport aux non-fumeuses, et que les femmes exposées durant la grossesse de leur mère avaient un risque relatif de 1,16 (95% IC : 1,01 - 1,32).

Par ailleurs, l'étude du WECARE réalisée sur 1 521 femmes ayant développé un cancer du sein asynchrone et 2 212 femmes atteintes d'un cancer du sein unilatéral a montré que les patientes consommant à la fois de l'alcool et du tabac durant leur premier cancer avait un risque de rechute 1,6 (95% IC : 1,24-2,11) fois supérieur aux femmes n'en consommant pas⁸⁰.

3.1.6.2 *Indice de masse corporelle*

Le surpoids (IMC>29) influencerait le risque de cancer du sein par une modification du taux d'œstrogènes⁸¹. En effet, l'obésité serait liée à des règles précoces et à une ménopause tardive, ce qui augmenterait l'exposition de la glande mammaire aux hormones sexuelles. Le risque de cancer du sein associé à l'IMC varie en fonction de l'état pré- ou post-ménopausé d'une femme. Une méta-analyse réalisée chez plus de 2,5 millions de femmes, dont 7 930 femmes pré-ménopausées atteintes de cancer du sein, montre que les femmes pré-ménopausées ont un risque de développer la maladie qui diminue de 8% par 5 kg/m² au-dessus du seuil de surpoids (RR=0,92, 95% IC : 0,88 – 0,97)⁸². Cet effet protecteur du surpoids s'inverse après la ménopause, comme le montre l'étude réalisée au Royaume-Uni chez 1,2 millions de femmes âgées de 50 à 64 incluant 45 037 femmes atteintes de cancer du sein⁸³ : les femmes post-ménopausées en surpoids présentent un risque de cancer du sein 1,29 supérieur aux femmes ayant un IMC compris entre 22,5 et 24,9 (95% IC : 1,22 – 1,36).

Il a été estimé que 38% à 50% des cancer du sein survenant dans les pays dits « développés » seraient évitables par une modification de nos habitudes de vie⁸⁴. Les facteurs environnementaux connus ne permettent donc pas à eux seuls d'expliquer la survenue de la maladie.

3.2 Facteurs de risque génétiques

L'histoire familiale est le deuxième facteur de risque le plus important après l'âge. Le risque relatif familial (RRF) d'une femme augmente en fonction du degré et du nombre d'apparentées atteintes d'un cancer du sein. Une femme présente un risque relatif familial de 1,80 (99% IC : 1,69 - 1,91) si sa mère a développé un cancer du sein, et ce risque est de 2,93 (99% IC : 2,36 - 3,64) si sa mère et sa sœur ont développé la maladie (Figure 9)⁸⁵.

L'augmentation du risque de cancer du sein est d'autant plus importante que l'âge et les antécédents familiaux sont pris en compte dans l'estimation de ce risque. Par exemple, il a été montré que le risque cumulé pour une femme de 50 ans sans apparentée de premier degré atteinte de cancer du sein est de 1,7% et que ce risque est respectivement de 3,7% et de 8,0% si cette femme a une ou deux apparentées du premier degré atteintes de cancer du sein⁸⁵.

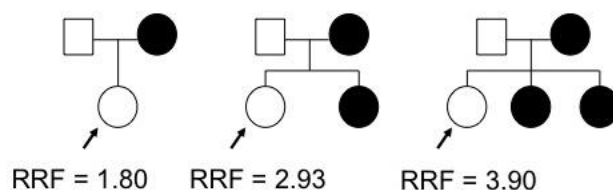


Figure 9 : Représentation du risque relatif familial de cancer du sein en fonction du nombre d'apparentées atteintes de la maladie.

RRF : Risque relatif familial. La flèche indique la femme pour qui le risque de cancer est estimé en fonction du nombre d'apparentées du premier degré ayant développé un cancer du sein.

Bien que le chirurgien français Paul Broca décrit dès 1866 dans le « Traité des Tumeurs » une famille dans laquelle une dizaine de cas de cancer du sein avaient été diagnostiqués, ce n'est qu'un siècle plus tard que des facteurs génétiques seront impliqués dans la prédisposition au cancer du sein. On estime aujourd'hui que 5 à 10% des cancers du sein sont dus à une prédisposition génétique, la majorité des cas étant considérés comme des cas « sporadiques » c'est à dire sans histoire familiale connue.

Depuis les années 1990, la liste des gènes, ou loci, impliqués dans la prédisposition génétique au cancer du sein n'a cessée de s'étoffer. On sait aujourd'hui classer les allèles associés à la maladie en trois catégories :

- Les allèles conférant un risque élevé, mis en évidence pour les gènes *BRCA1* (*BReast CAncer susceptibility 1*) et *BRCA2* (*BReast CAncer susceptibility 2*), ou bien pour les gènes impliqués

dans les syndromes de Li-Fraumeni (*TP53*), Peutz-Jegher (*STK11*), Cowden (*PTEN*) ou les cancers gastriques familiaux (*CDH1*).

- Les allèles conférant un risque modéré, associés notamment aux gènes *ATM*, *CHEK2*, *RINT1*, *XRCC2* et mis en évidence par le reséquençage de ces gènes dans des familles à haut-risque de cancer du sein et dans des études cas-témoins multicentriques.
- Les allèles conférant un risque faible, associés à des variants fréquents dans la population et identifiés par les études pangénomiques, comme les études du *Breast Cancer Association Consortium* (BCAC).

Nous allons dans ce chapitre détailler les caractéristiques de ces trois catégories d'allèles associés à la prédisposition au cancer du sein.

3.2.1 Allèles associés à un risque élevé de cancer du sein

3.2.1.1 Gènes *BRCA1* et *BRCA2*

3.2.1.1.1 Découverte des gènes

Dans les années 1990, les études de liaison génétique sont réalisées dans des familles présentant plusieurs cas de cancer du sein et de l'ovaire afin d'identifier des gènes responsables du cancer du sein. Les analyses de liaison utilisant la méthode des lod-scores (LOD) sont basées sur la co-ségrégation, au cours des générations, des allèles responsables de la maladie dans les familles. Le LOD est calculé en définissant soit un modèle de transmission de la maladie récessif ou dominant, la fréquence de l'allèle à risque dans la famille et la pénétrance de la mutation. L'âge peut également être pris en compte dans l'estimation du LOD. Par cette méthode, deux gènes majeurs de prédisposition, *BReast CAncer susceptibility 1* (*BRCA1*) localisé sur le locus 17q21 et *BReast CAncer susceptibility 2* (*BRCA2*) localisé sur le locus 13q13 sont mis en évidence. En 1990, le premier gène de prédisposition *BRCA1* est mis en évidence par l'équipe de Marie-Claire King⁸⁶ à partir de 23 familles présentant plusieurs cas de cancer du sein et/ou de l'ovaire. Quatre ans plus tard ce locus est cloné afin d'en définir la séquence nucléotidique précise⁸⁷. *BRCA1* est un gène mesurant 81,2kb et est composé de 24 exons dont 22 codant pour une protéine de 1863 acides aminés (AA).

La même année où *BRCA1* a été cloné, un deuxième locus situé en 13q12-13 est mis en évidence par Wooster et coll.. Le clonage de ce locus révélera le gène *BRCA2* mesurant 84kb et composé de 27 exons dont 26 codant pour une protéine de 3418 AA.

3.2.1.1.2 Structure des protéines *BRCA1* et *BRCA2*

D'un point de vue structurel, les protéines *BRCA1* et *BRCA2* présentent peu de similitudes. La protéine *BRCA1* est composée de plusieurs domaines ayant chacun un rôle bien déterminé : le

domaine RING permet la liaison avec la protéine BARD1 pour activer les fonctions d'ubiquitine ligase de BRCA1 ; les deux domaines BRCT (*BRCA1 C-Terminal*) permettent l'interaction de la protéine avec des partenaires impliqués dans les mêmes voies biologiques ; les deux domaines NLS (*Nuclear Localization Sequence*) assurent la localisation de la protéine au noyau et le domaine SCD (séquence de sérines et de thréonines) est préférentiellement phosphorylé par ATM⁸⁸ (Figure 10.A).

La protéine BRCA2, plus longue que BRCA1 possède un domaine BRCT plusieurs fois répété et deux domaines NLS⁸⁸ (Figure 10.B).

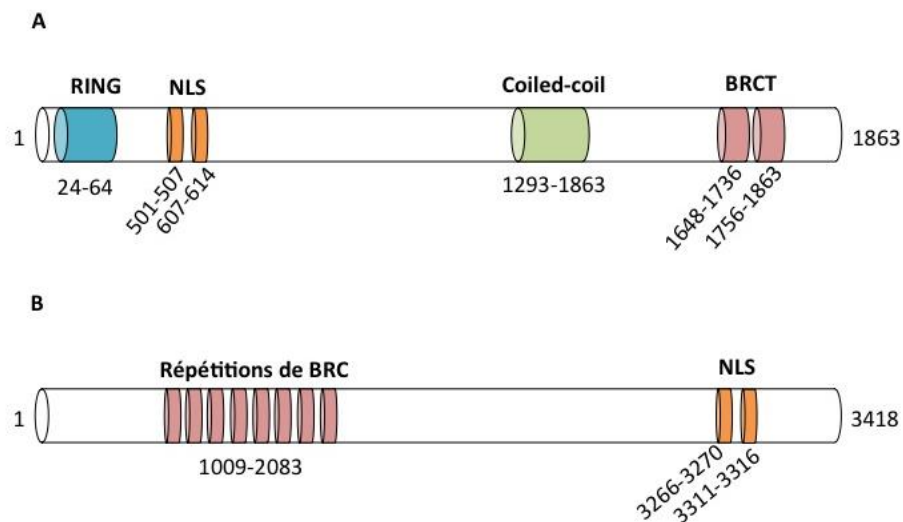


Figure 10 : Domaines fonctionnels des protéines BRCA1 et BRCA2. A. Domaines fonctionnels de BRCA1.

A. Domaines fonctionnels de BRCA1 : le domaine RING interagit avec la protéine BARD1 ; NLS : *nuclear localization sequence* permet la localisation de BRCA1 au noyau ; le domaine coiled-coil permet la liaison avec PALB2 ; BRCT : *BRCA1 C-terminal* interagit avec les protéines ABRAXAS, CtIP et BRIP1 ; **B. Domaines fonctionnels de BRCA2.** BRC : domaines BRCA2 permettent la liaison avec RAD51 ; NLS : *nuclear localization sequence* permet la localisation de BRCA2 au noyau (d'après Roy et coll., Nature review, 2012)⁸⁹.

3.2.1.1.3 Fonction des protéines BRCA1 et BRCA2

La protéine BRCA1 est impliquée dans plusieurs voies biologiques permettant de maintenir l'intégrité du génome en contrôlant par exemple la prolifération cellulaire ou en intervenant dans la réparation de l'ADN.

BRCA1 contrôle la prolifération cellulaire des tissus hormono-dépendants en se liant au récepteur de l'œstrogène RE α pour inhiber, par exemple, la transcription du gène *VEGF* (*Vascular Endothelial Growth Factor*) responsable de l'angiogenèse des endothéliums⁹⁰.

Une autre fonction de BRCA1 est de contrôler le point de passage du cycle cellulaire, notamment des phases G2 à M. En cas de dommage à l'ADN, BRCA1 s'associe à CHEK1 pour inhiber l'action de cycline B afin de bloquer le passage des cellules en phase de mitose et enclencher la réparation des dommages⁹¹. BRCA1 joue également un rôle essentiel dans la réparation des cassures doubles brins de l'ADN ou des liaisons inter-brins par recombinaison homologue (RH). Grâce à son domaine de liaison à l'ADN et aux domaines BRCT, BRCA1 peut détecter les cassures doubles-brins de l'ADN et

recruter les protéines nécessaires à la réparation par RH (Figure 11). BRCA1 a également une fonction d'ubiquitine-ligase via son domaine BRCT qui facilite le transfert des ubiquitines aux protéines cibles. L'interaction de BARD1 avec BRCA1 stimule l'activité d'ubiquitine-ligase de BRCA1 pendant la réparation des dommages à l'ADN et pendant la réplication de l'ADN⁹².

La protéine BRCA2 est également impliquée dans la réparation des cassures doubles-brin de l'ADN en maintenant la protéine RAD51 inactivée. L'interaction du domaine BRC de BRCA2 avec RAD51 maintient inactive cette dernière⁹³ et régule sa fixation avec BRCA1.

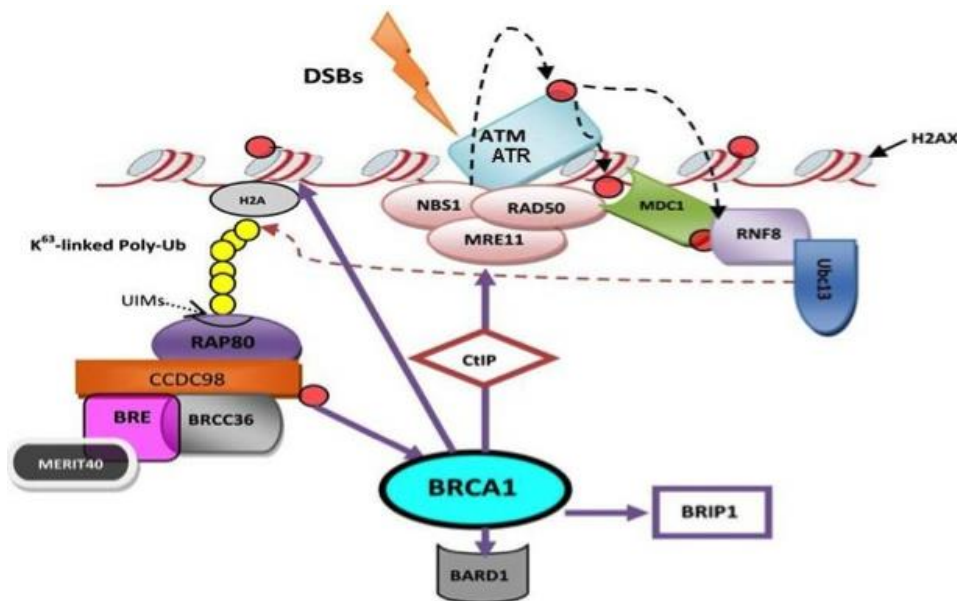


Figure 11 : Réparation des cassures double-brin de l'ADN.

Via son domaine BRCT, BRCA1 interagit directement avec les protéines BRIP1, BARD1, CCDC98 (ABRAXAS) et CtIP afin de recruter les protéines essentielles à la réparation des dommages à l'ADN.

Les gènes *BRCA1* et *BRCA2*, de par leur rôle dans la maintenance de l'intégrité du génome sont considérés comme des gènes suppresseurs de tumeur. Une perte d'expression de ces gènes conduit au développement tumoral.

3.2.1.1.4 Les allèles de prédisposition au cancer du sein

La prévalence des allèles délétères de *BRCA1* et *BRCA2* dans la population générale est faible ; ces allèles sont identifiés dans moins de 0.05% à 0.5% de la population générale⁹⁴. Des milliers de mutations ont été recensées, majoritairement des mutations pertes de fonction réparties sur tout le gène. Les bases françaises Universal Mutation database *BRCA1* et *BRCA2* (UMB-*BRCA1* et UMD-*BRCA2*) recensent les différents variants identifiés chez des cas indexés vus en consultation de génétique et testés pour *BRCA1* et *BRCA2*⁹⁵. Avant d'être incrémentés dans la base, tous les variants sont validés et stratifiés en cinq classes : causal, probablement causal, variant de signification inconnu (VSI), probablement bénin et bénin. A ce jour 897 variants causaux de *BRCA1* identifiés chez 5 637 familles, et 968 variants causaux de *BRCA2* identifiés chez 4 185 familles sont répertoriés dans

les bases françaises UMD-BRCA1 et UMD-BRCA2. On observe cependant quelques points chauds (« hotspot »), comme le montre une étude dans la population française chez qui des mutations causales sont plus fréquemment trouvées dans les exons 2, 5 et 20 de *BRCA1* et dans les exons 8, 13 et 23 de *BRCA2*⁹⁵. Il existe également quelques mutations fondatrices dans la population juive Ashkénaze originaire d'Europe de l'Est (mutation *BRCA1* : c.185_186del⁹⁶ ; mutation *BRCA2* : c.6174del⁹⁷), dans la population québécoise issue de l'immigration française (mutations *BRCA1* : c.4446C>T, c.2834_2836delinsC, c.3768dup et c.2598C>A ; mutation *BRCA2* : c.8765_8767del, c.6085G>T, c.2816_2817insA and c.6503_6504del)^{98, 99} ou encore dans la population polonaise (mutations *BRCA1* : c.5383dup, c.61C>G et c.4153del¹⁰⁰ ; mutations *BRCA2* : c.9631del et c.6886_6887del)¹⁰¹.

Les premiers risques absolus (ou pénétrance) pour les porteurs d'une mutation dans l'un de ces deux gènes ont été estimés à partir de familles sélectionnées pour leur histoire familiale de cancer du sein ou de l'ovaire¹⁰². Dans ces familles, le risque de développer un cancer du sein pour les porteurs d'une mutation dans l'un ou l'autre des deux gènes a été estimé à 74% pour une femme âgée de 70 ans¹⁰². Cependant en raison du biais de sélection des familles on sait aujourd'hui que ces risques ont été surestimés. Différentes méta-analyses réalisées dans les années 2000 à partir de cas non-sélectionnés sur leur histoire familiale ont réévalué le risque de cancer du sein associé aux mutations dans ces gènes. D'après la méta-analyse de Antoniou et coll. portant sur plus de 8 000 cas familiaux non sélectionnés sur la présence d'une mutation *BRCA1* ou *BRCA2*, le risque cumulé à 70 ans pour les porteurs *BRCA1* est estimé à 65% et à 39% pour les porteurs *BRCA2*¹⁰³. Bien que les femmes soient le plus affectées par la maladie, les mutations de *BRCA1* et *BRCA2* augmentent également le risque de cancer du sein chez l'homme ; le risque cumulé à 70 ans est de 1,2% pour les porteurs d'une mutation *BRCA1* et de 6,8% pour les porteurs d'une mutation *BRCA2*¹⁰⁴.

En raison des risques élevés de cancer du sein et de l'ovaire pour les porteuses d'une mutation dans les gènes *BRCA1* et *BRCA2*, un test génétique est proposé aux personnes malades (cas index) et non malades dans les familles hautement prédisposées vues en consultation d'oncogénétique pour rechercher la présence d'une mutation dans l'un de ces gènes. Un test positif implique des recommandations et un suivi bien particulier pour ces patientes et leur famille, consistant en un examen clinique mammaire annuel avant 30 ans et après 65 ans et en une imagerie par résonance magnétique (IRM) et une mammographie entre 30 et 65 ans. Une réduction mammaire annexielle peut être proposée à 40 ans pour les porteuses *BRCA1* et à 45 ans pour les porteuses *BRCA2* (recommandations françaises 2017 de l'INCa).

3.2.1.2 Le gène *PALB2*

Le gène *PALB2*, mis en évidence chez des enfants atteints d'anémie de Fanconi, est également associé à une prédisposition génétique au cancer du sein^{105, 106}. Les premières études réalisées dans des familles de cas multiples de cancer du sein, sans mutation *BRCA1* ou *BRCA2*, estiment que les mutations délétères de *PALB2* confèrent un risque de 2 à 6,1 fois supérieur à celui de non-porteurs^{106, 107}. Des mutations fondatrices de ce gène ont également été mises en évidence dans la population canadienne (c.2323C>T ; p.Gln775X)¹⁰⁸ et finlandaise (c.1592del ; p.Leu531Cysfs30)¹⁰⁵.

Les variants de *PALB2* ont d'abord été associés à un risque intermédiaire de cancer du sein jusqu'aux travaux du *PALB2 Interest group*¹⁰⁹. D'après cette étude, les porteurs de mutations tronquantes de *PALB2* ont un risque 9 fois supérieur à celui de la population générale et un risque absolu à 70 ans qui varie de 33% pour les femmes sans apparentées atteintes de cancer du sein à 58% pour les femmes ayant une ou deux apparentées atteintes¹⁰⁹. Ces résultats suggèrent que des variants délétères de *PALB2* seraient associés à un risque de cancer du sein proche du risque estimé pour les porteurs d'une mutation de *BRCA2*. Depuis cette étude, les patients porteurs d'une mutation délétère dans ce gène reçoivent des recommandations et un suivi plus adaptés. Aux Etats-Unis, le *National Comprehensive Cancer Network* (NCCN) recommande une mammographie annuelle pour les femmes de 30 ans et plus, une IRM peut également être réalisée et une mastectomie prophylactique peut être prescrite selon l'histoire familiale. En France, le Groupe Génétique et Cancer d'Unicancer (<http://www.unicancer.fr/la-recherche-unicancer/les-groupes-transversaux/groupe-genetique-cancer-ggc>) regroupant des cliniciens, des biologistes, des chercheurs et des épidémiologistes a établi des recommandations similaires : une mammographie, un IRM ou un sonogramme du sein (échographie) sont prescrits pour les femmes de 30 à 65 ans et une mastectomie bilatérale peut également être proposée.

3.2.1.3 Gènes associés à des syndromes

En plus des gènes *BRCA1*, *BRCA2* et *PALB2*, les gènes *TP53*, *PTEN* (*Phosphatase and Tensin homology deleted on chromosom ten*), *STK11* (*Serine/Threonine kinase 11*) et *CDH1* (*Cadherin 1*) impliqués dans différents syndromes confèrent un risque accru de cancer du sein chez les patients atteints par ces syndromes.

3.2.1.3.1 *TP53* et syndrome de Li-Fraumeni

Le syndrome de Li-Fraumeni est un syndrome héréditaire autosomique dominant causé par des mutations délétères dans le gène *TP53*¹¹⁰. Les tumeurs les plus caractéristiques associées à ce syndrome sont les ostéosarcomes, les sarcomes des tissus mous, les leucémies ou lymphomes, les

tumeurs cérébrales et également le cancer du sein. En effet, le risque de cancer du sein est de 15% à 15 ans et de 80% à 50 ans en présence d'une mutation de *TP53*¹¹¹.

Selon les critères de Chompret, une famille est suspectée comme affectée par ce syndrome si ¹¹²:

- un patient est atteint d'un cancer appartenant au spectre tumoral lié au syndrome de Li-Fraumeni avant 46 ans (sarcome, ostéosarcome, cerveau, cancer du sein pré-ménopause, carcinome des glandes surrénales, leucémie) et qu'un apparenté du premier ou second degré présente soit une tumeur liée à ce syndrome avant 56 ans soit des cancers multiples ;
- un patient est atteint de tumeurs multiples (hors cancer du sein multiple) dont deux tumeurs sont liées au spectre tumoral du syndrome de Li-Fraumeni, un cancer étend survenu avant 46 ans ;
- un enfant atteint d'un corticosurréalome.

3.2.1.3.2 *PTEN* et le syndrome de Cowden

Le syndrome de Cowden est un syndrome rare (1/200 000) à transmission autosomique dominante lié à des mutations du gène *PTEN*¹¹³. La maladie se caractérise par des hamartomes (malformation des tissus d'origine embryonnaire) de la peau et de la thyroïde, mais également par une macrocéphalie, une ataxie cérébelleuse et un risque accru de tumeurs mammaires, de l'endomètre ou de la thyroïde. Le risque cumulé de cancer du sein chez les personnes atteintes de ce syndrome à 70 ans est de 77% (95% IC : 59%-91%)¹¹⁴.

3.2.1.3.3 *STK11* et le syndrome de Peutz-Jeghers

Le syndrome de Peutz-Jeghers est une maladie gastro-intestinale héréditaire rare apparaissant durant l'enfance ou l'adolescence et qui se caractérise par le développement de polypes hamartomateux au niveau du tube digestif¹¹⁵. Ce syndrome est causé par des mutations délétères du gène *STK11*. Les femmes atteintes de ce syndrome ont un risque accru de cancer du sein, avec un risque cumulé de 31% (95% IC : 18-50) à 50 ans¹¹⁶.

3.2.1.3.4 *CDH1* et les formes héréditaires de cancers gastriques

Les formes familiales de cancer gastriques sont liées à des mutations dans le gène *CDH1*¹¹⁷. Ces familles présentent également un risque accru de cancer du sein et particulièrement de cancer lobulaire avec un risque cumulé de 42% à 80 ans¹¹⁸.

Les patients affectés par les formes familiales de cancer gastrique, par les syndromes de Cowden et de Li-Fraumeni ont un suivi adapté dû au fort risque de développer un cancer du sein. En effet, les patients atteints d'un syndrome de Cowden ou d'une forme familiale de cancer gastrique ont une surveillance mammaire de 30 à 65 ans consistant en une mammographie, une IRM ou un sonogramme annuel. Pour les patients atteints du syndrome de Li-Fraumeni, la surveillance

mammaire par IRM ou sonogramme commence à l'âge 20 ans. Une mastectomie prophylactique pour l'ensemble de ces patients peut également être proposée.

3.2.2 Allèles associés à un risque faible de cancer du sein

La mise en évidence des allèles conférant un risque faiblement accru de cancer du sein repose sur l'hypothèse du modèle polygénique. Le développement de la maladie ne serait pas dû à une seule mutation mais à un ensemble de facteurs, notamment à des variants relativement fréquents dans la population ; dans cette hypothèse, chaque variant pris individuellement serait associé à une augmentation du risque minime, mais l'effet cumulé de ces variants augmenterait fortement le risque de développer un cancer du sein.

Grâce aux progrès technologiques, en particulier à l'avènement des puces à ADN permettant le génotypage simultané de dizaine de milliers de SNPs (*Single Nucleotide Polymorphisms*) et à l'accès à de grandes études cas-témoins ou à des cohortes prospectives menées le plus souvent dans la population générale, un grand nombre de variants fréquents dans la population ont ainsi pu être associés individuellement à une augmentation faible du risque du cancer du sein (OR de l'ordre de 1,2 - 1,4).

Les premières études d'association de type cas-témoins ont ciblé des SNPs candidats ayant une fréquence de l'allèle mineur (FAM) d'au moins 5% et localisés dans des gènes impliqués dans la réparation des dommages à l'ADN, le cycle cellulaire ou encore la carcinogenèse. En 2004 et 2005, deux études portant respectivement sur 6 000 et 1 355 participants^{119, 120} mettent en évidence un SNP dans le gène *CASP8* associé à une diminution du risque de cancer du sein pour les porteurs du variant. Ces résultats seront par la suite validés dans une étude plus large du *Breast Cancer Association Consortium* (BCAC) impliquant 16 423 cas and 17 109 témoins (OR=0.89 (95%CI: 0.85–0.94))¹²¹.

Grâce au succès de ces études, au développement de nouvelles générations de puces à ADN, pouvant aujourd'hui cibler plusieurs millions de SNPs, et à la diminution des coûts de génotypage, de plus larges études ont vu le jour. Dès 2007, les approches « SNPs ou gènes candidats » ont été remplacées par des études pangénomiques ou « *Genome Wide-Association Studies* » (GWAS). Ces études ont permis d'identifier plus d'une centaine de SNPs associés au risque de cancer du sein. En 2013, plusieurs études internationales s'intéressant aux cancers du sein (chez les porteurs et les non-porteurs d'une mutation dans *BRCA1* ou *BRCA2*), de l'ovaire et de la prostate forment le consortium *Collaborative Oncological Gene-environment Study* (COGS)¹²² et élabore la puce iCOGS (Illumina), ciblant plus de 211 000 SNPs chez plus de 200 000 participants. Ces travaux ont permis de mettre en évidence des SNPs localisés dans plus de 100 loci, notamment au niveau des gènes *FGFR2*, *TNRC9*,

MAP3K1, *COX11*, *NEK10*, *RAD51L1*, *11q13*, *ZMIZ1*, *CDKN2A/B*, *ZNF35* et *ANKRD16* ainsi que dans les régions intergéniques en 2q35, 6q25 et 5q12¹²³⁻¹²⁵. Là encore, ces SNPS sont associés à une augmentation faible du risque de cancer du sein, de l'ordre de 1,2 à 1,5 pour les cas dits sporadiques. D'autre part études réalisées par le consortium CIMBA, auquel contribue l'étude française GEMO (Gènes modificateurs des risques tumoraux chez les sujets portant une mutation de *BRCA1* ou *BRCA2*) ont montré que certains loci associés au risque de cancer du sein modifient particulièrement le risque de développer la maladie chez les porteurs d'une mutation dans les gènes *BRCA1* et *BRCA2*. Par exemple, être porteur de l'allèle mineur des SNPs rs11249433 (*EMPB1*)¹²⁶, rs10995190 (*ZNF365*)¹²⁷, rs889312 (*MAP3K1*)¹²⁸ ou rs3817198 (*LSP1*)¹²⁹ augmente le risque de développer la maladie chez les porteurs *BRCA2*, alors qu'être porteur de l'allèle mineur des SNPs rs13387042 (*IGFPB5*) ou rs3803662 (*TNRC9*) augmente le risque de cancer du sein à la fois chez les porteurs *BRCA1* et les porteurs *BRCA2*.

Bien que l'identification des SNPs permette d'expliquer environ 28% de l'héritabilité, environ la moitié de la prédisposition familiale restait inexpliquée. On estime aujourd'hui que des milliers de SNPs impliqués dans la prédisposition génétique au cancer du sein restent à identifier¹³⁰. Des analyses menées par les consortia BCAC et CIMBA sont actuellement en cours pour évaluer l'implication de nouveaux SNPs dans différents groupes de population. Ces nouvelles analyses exploitent en plus des données générées avec la puce iCOGS les informations apportées par une nouvelle puce ciblant plus de 600 000 SNPs (puce OncoArray d'Illumina)¹³¹.

A l'heure actuelle, une réflexion sur l'intérêt clinique d'incorporer un score de risque polygénique (PRS pour *Polygenic Risk Score*) dans les modèles de prédiction commence à émerger.

3.2.3 Allèles associés à un risque modéré de cancer du sein

L'une des démarches actuelles pour identifier de nouveaux gènes de prédisposition est le re-séquençage de gènes candidats dans des études familiales et/ou cas-témoins pour mettre en évidence des facteurs génétiques qui pourraient contribuer à l'apparition du cancer du sein. En général, les gènes sont sélectionnés dans ces études selon leur implication dans différentes voies biologiques liées par exemple à la réparation de l'ADN ou au cycle cellulaire (Figure 12). Certains sont parfois aussi mis en lumière par séquençage de l'exome de cas familiaux, comme par exemple les gènes *XRCC2* et *RINT1*^{132, 133}.

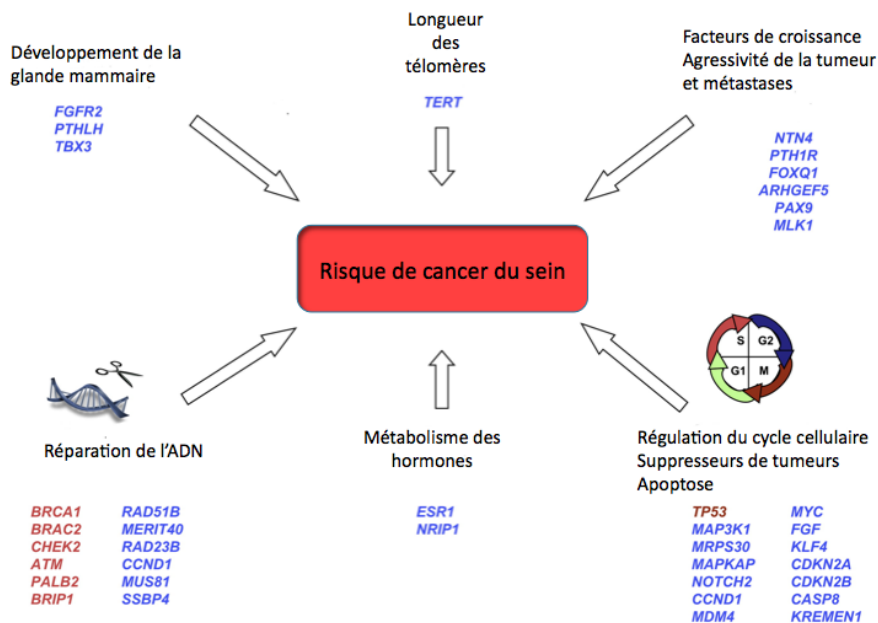


Figure 12 : Mécanismes potentiellement impliqués dans le risque de cancer du sein et les gènes connus qui y sont associés.

Les gènes en rouge correspondent aux allèles de prédisposition à forte et moyenne pénétrance, alors que les gènes en bleu correspondent aux allèles à faible pénétrance. (Adapté de Ghossaini M et al, 2013)¹³⁴.

3.2.3.1 ATM

En 2006, Renwick et coll. publient une étude dans laquelle la séquence codante du gène *ATM* a été entièrement reséquencée chez 443 cas familiaux de cancer du sein et 521 témoins, et confirment la contribution des variants tronquants (perte de fonction) d'*ATM* dans la prédisposition à la maladie. Le risque estimé pour ces variants est environ 2 fois supérieur à celui de non-porteur (RR = 2,4 (95% IC : 1,51-3,78))¹³⁵. Ce résultats n'était pas une « découverte » puisqu'en 1987, Swift et coll. mettaient déjà en évidence que les femmes porteuses d'une mutation dans ce gène ont un risque de cancer du sein 3 fois plus élevé que les femmes non-porteuses de la mutation.

Deux ans plus tard, Tavtigian et coll. réalisent une analyse groupée à partir de données publiées provenant de sept études, regroupant 1 544 cas de cancer du sein et 1 224 témoins, et du séquençage d'*ATM* dans la ressource multicentrique du Breast Cancer Family Registry (BCFR) impliquant plus de 1200 cas précoces de cancer du sein (<45 ans) et 1200 témoins d'Australie, du Canada et des USA¹³⁶. Cette étude a ainsi évalué l'implication dans la prédisposition au cancer du sein de variants rares tronquants ou faux-sens situés dans la séquence codante d'*ATM*. L'impact des variants faux-sens sur la protéine a été évalué en fonction de la conservation de l'acide aminé au cours de l'évolution grâce à l'outil *in silico* Align-GVGD classant les variants faux-sens en 7 grades, du grade le moins délétère C0 (effet probablement bénin) au grade le plus délétère C65 (effet

probablement délétère), avec cinq classes associés à des effets intermédiaires C15, C25, C35, C45, C55. La distribution des variants rares tronquants et faux-sens a ensuite été comparée chez les cas et les témoins afin d'estimer les risques (Odd Ratio, OR) de cancer sein associé à ces variants. Les résultats obtenus pour les variants tronquants à partir du séquençage d'*ATM* dans l'étude du BCFR et de l'analyse groupée confirment les résultats de Renwick et *coll.* (OR mutations tronquantes = 2,08 (95%IC : 1,03-4,21)) (Tableau 5). Cette étude montre également que des variants faux-sens dans le gène augmentent significativement le risque de cancer du sein et que cette augmentation du risque de cancer du sein pourrait être plus élevée que l'augmentation du risque associé aux mutations tronquantes (Tableau 6). Cette dernière étude montre donc l'intérêt de prendre en compte tous les types de variants dans les analyses cas-témoins. L'application de la même stratégie dans l'étude du BCFR a d'ailleurs permis d'évaluer la contribution des variants codants rares de plusieurs autres gènes candidats impliqués dans la réparation des cassures doubles-brins de l'ADN. ; nous allons détailler ces gènes dans le paragraphe 3.2.3.2.

Tableau 5 : Risque de cancer du sein estimé par l'analyse groupée chez les porteurs d'un variant tronquant d'*ATM*.

	Cas	Témoins	OR ajusté (95% IC)*
Séquençage d'<i>ATM</i> dans l'étude du BCFR			
Non-HetAT	2 505	2 235	ref
Porteurs d'un VT	26	10	2.08 (1.03-4.21)
Etude groupée			
Non-HetAT	4 076	2 389	ref
Porteurs d'un VT	36	10	2.08 (1.03-4.21)

*Odd ratio ajusté sur l'origine ethnique et le centre d'inclusion. VT : variants tronquants (d'après Tavitigan et al. 2009)¹³⁶

Tableau 6 : Risque de cancer du sein estimé chez les porteurs d'une substitution faux-sens d'*ATM*

	Cas	Témoins	OR ajusté (95% IC)*
Non-HetAT	1 788	1 717	ref
C0	86	89	0.93 (0.68-1.26)
C15	34	29	1.13 (0.68-1.86)
C25	9	7	1.24 (0.46-3.33)
C35	0	1	-
C45	1	0	-
C55	5	4	1.20 (0.32-4.49)
C65	25	5	6.00 (2.09-1.26)

*Odd ratio ajusté sur l'origine ethnique et le centre d'inclusion. L'effet des variants faux-sens a été évalué avec l'outil de prédiction Align-GVGD en 7 grades, du grade le moins délétère (C0) au plus délétère (C65) (d'après Tavitigan et al. 2009)¹³⁶.

Une autre étude de type cas-témoins portant sur le gène *ATM* réalisée dans la population australienne confirme les résultats publiés par Renwick et coll. et Tavtigian et coll.. Cette étude a été menée dans un échantillon indépendant de 2 570 cas et 1 448 témoins; elle rapporte une augmentation du risque (OR = 2,55 (95%IC : 0.54-12.0)) pour les porteurs de variants faux-sens classés C65 ou C55¹³⁷. Dans cette dernière étude les auteurs se sont intéressés plus spécifiquement au variant c.7271T>G (p.Val2424Gly) prédit comme délétère et souvent identifié dans des familles australiennes à haut risque de cancer du sein (prévalence du variant : 0,2% de la population australienne¹³⁸ et <0.1% dans la population Française (Girard, Lesueur et coll. Etude GENESIS, communication personnelle)). Les porteurs de ce variant ont un risque 8 fois plus élevé que les non-porteurs d'une mutation d'*ATM* (95% IC : 2,3-27,4).

L'ensemble de ces résultats confirment donc l'implication d'*ATM* dans la prédisposition au cancer du sein.

3.2.3.2 *Autres gènes de la réparation de l'ADN ayant été associés à un risque intermédiaire de cancer du sein*

3.2.3.2.1 *CHEK2 et les gènes du complexe MRN*

Le reséquençage des régions codantes du gène *CHEK2* et des gènes du complexe MRN (*MRE11*, *RAD50* et *NBN*) dans l'étude du BCFR^{139, 140} a montré, comme pour *ATM*, que les variants tronquants mais également les variants faux-sens conféraient une augmentation du risque de cancer du sein (OR_{CHEK2_tronquant} = 6.18 (95%IC : 1.76-21.8) versus OR_{CHEK2_faux-sens} = 2.2 (95%IC : 1.2-4.01); OR_{MRN_tronquants} = 2.61 (0.67-10.1) versus OR_{MRN_faux-sens} = 3.17 (95%IC : 1.17-8.59)).

Concernant le gène *CHEK2*, il existe comme pour *BRCA1* et *BRCA2*, des mutations fondatrices impliquées dans le risque de cancer du sein et présentent dans plus de 1% des populations d'Europe du nord (c.1100delC)¹⁴¹⁻¹⁴³ et dans la population Slave (p.Ile157Thr). Les porteurs de la mutation tronquante c.1100delC présentent un risque deux fois supérieur¹⁴⁴ à celui de non-porteurs alors que la mutation faux-sens p.Ile157Thr augmenterait le risque de 1,4¹⁴⁵.

3.2.3.2.2 *RAD51*

La protéine RAD51 joue un rôle essentiel dans la recombinaison homologue de l'ADN. Plusieurs études ont montré l'implication des paralogues de *RAD51*, *RAD51C*¹⁴⁶, *RAD51D*¹⁴⁷ et *XRCC2*¹³², dans la prédisposition au cancer du sein, mais trois études ayant reséquencées RAD51 dans des familles finlandaises et australiennes à haut-risque de cancer du sein ou de l'ovaire n'ont pas montré d'association du gène avec le risque de cancer du sein¹⁴⁸⁻¹⁵⁰. Le reséquençage du gène *RAD51* chez 1 330 cas et 1 123 témoins de l'étude du BCFR a confirmé ses résultats¹⁵¹.

3.2.3.2.3 *XRCC2 et RINT1, : deux nouveaux gènes de prédisposition au cancer du sein ?*

Plus récemment, les projets de séquençage de l'exome de femmes hautement prédisposées au cancer du sein ont mis en évidence le rôle de nouveaux gènes, impliqués comme les gènes cités plus haut dans des voies de la réparation de l'ADN. Une étude internationale a d'abord mis en évidence la présence de variants probablement délétères dans le gène *XRCC2* dans un groupe de femmes prédisposées au cancer du sein puis l'implication de ce gène dans la maladie a ensuite été confirmée dans l'étude cas-témoins du BCFR¹³². Une deuxième étude de type cas-témoins réalisée chez 3548 cas et 1435 témoins n'a cependant pas confirmé ce gène comme gène de prédisposition au cancer du sein¹⁵². Plus récemment, la prévalence des variants rares tronquants (perte de fonction) et des variants faux-sens potentiellement délétères a été par la suite évaluée dans un échantillon comptant plus de 54 000 femmes ayant reçu un test génétique dans le cadre d'une prédisposition au cancer du sein et 38 000 témoins de *l'Exome Aggregation Consortium*¹⁵³ ; les résultats de cette étude cas-témoins ne confirment pas les résultats décrits par Park et coll.

L'approche en deux temps (analyse de l'exome de cas familiaux de cancer du sein, puis validation d'un gène candidat dans une étude cas-témoins) utilisée dans l'étude du BCFR portant sur *XRCC2* a également permis de mettre en lumière le gène *RINT1*. Le reséquençage de l'exome de 89 femmes atteintes d'un cancer du sein à un jeune âge a mis en évidence trois variants délétères dans ce gène dans trois familles différentes¹³³. Le reséquençage de *RINT1* dans la population de l'étude BCFR a montré une association significative entre les variants rares de *RINT1* et le risque de cancer du sein (OR, 3.24; 95% ICI : 1,29–8,17). Deux ans plus tard, Li et coll. séquencent les régions codantes de *RINT1* chez 2 024 cas de cancer du sein de patients non-porteurs de mutations dans les gènes *BRCA1*, *BRCA2* et *PALB2* et chez 1 886 témoins¹⁵⁴. Dans cette dernière étude, la proportion de variants rares tronquants ou faux-sens identifiés chez les cas est similaire à la proportion mis en évidence chez les témoins, suggérant que *RINT1* n'est pas impliqué dans la prédisposition au cancer du sein dans la population étudiée.

3.2.3.2.4 *ABRAXAS*

Dans le cadre de mon projet de Master j'ai participé au projet du BCFR et ai reséquencé le gène *ABRAXAS (FAM175A)* chez 1330 cas précoces de cancer du sein et 1123 témoins pour déterminer si des variants rares (<1%) de ce gène contribuaient à la prédisposition génétique au cancer du sein. Le choix de ce gène candidat reposait sur le rôle d'ABRAXAS dans le recrutement de BRCA1 lors de la réparation des cassures doubles-brin de l'ADN (Figure 11). En plus de deux SNPs, seize variants rares, dont un variant « inframe », ont été identifiés dans la population du BCFR mais aucun de ces variants

n'a été associé au cancer du sein¹⁵⁵. Néanmoins, les essais d'immunofluorescences réalisés pour déterminer l'impact des variants faux-sens classés comme dommageables pour la protéine par l'outil de prédiction Align-GVGD (p.Thr141Ile et p.Gly39Val) montrent une diminution du nombre de foci formés par ABRAXAS au niveau des histones γ H2AX, suggérant que ces deux variants affectent la réparation des cassures doubles-brins de l'ADN. Les résultats de ces travaux qui ont fait l'objet d'une publication suggèrent que seule une petite proportion des cancers du sein héréditaire pourrait être due à des variants rares d'ABRAXAS (ANNEXE 1).

3.2.3.3 BRIP1

Le gène *BRIP1* est impliqué dans l'anémie de Fancomie¹⁵⁶ et dans la réparation des cassures doubles-brins de l'ADN. En 2006, Seal et coll. ont séquencé le gène chez 1 212 femmes atteintes de cancer du sein et 2 012 témoins et ont montré que les porteurs d'un variant tronquant de *BRIP1* avaient un risque 2 fois supérieur à celui de non-porteurs (95% IC : 1,2-3,2)¹⁵⁷. Dix ans plus tard, Easton et coll. génotypent chez 48 143 cas et 43 608 témoins inclus dans le consortium BCAC la mutation tronquante p.Arg798* et 10 variants faux-sens¹⁵⁸, et séquence également les régions codantes de *BRIP1* chez 13 213 cas et 5 242 témoins. Les résultats suggèrent que les variants de *BRIP1* ne sont pas impliqués dans la prédisposition au cancer du sein ($OR_{\text{variants_tronquants}}=0,9$ (95% IC : 0,48-1,7) ; $OR_{\text{variants_faux-sens}}=1,08$ (95% IC : 0,95-1,24).

3.2.3.4 Gènes impliqués dans le syndrome de Lynch

Le syndrome de Lynch également appelé cancer colorectal héréditaire sans polypose (HNPCC pour « hereditary non-polyposis colorectal cancer ») est la forme héréditaire de cancer colorectal la plus connue. Le syndrome de Lynch est soupçonné dans une famille lorsqu'un patient de moins de 50 ans développe un HNPCC ou lorsque d'autres membres de la famille ont été atteints entre 40 et 50 ans. Ce syndrome rare (3% des cancers colorectaux) se développe à l'âge adulte et est causé par des mutations dans les gènes de la réparation des bases non-appariées (« gènes MMR » pour MisMatch Repair) : *MLH1*, *MSH2*, *MSH6* et *PMS2*¹⁵⁹. Les personnes atteintes de ce syndrome ont également un risque augmenté de cancers de l'ovaire, des voies urinaires et biliaires, de l'intestin grêle et de la peau. En revanche, les évidences d'une augmentation du risque de cancer du sein ne sont pas probantes et ce sur-risque n'est pas pris en compte dans les recommandations de l'INCa de 2009.

L'ensemble des études d'association ont permis d'identifier de nouveaux gènes de prédisposition conférant un risque relatif pour les porteurs 2 à 5 fois supérieur à celui de la population générale et

expliquent environ 3% du risque familial de cancer du sein (Figure 13). *ATM*, *BRIP1*, *BARD1* et *CHEK2* entre autre, sont maintenant inclus dans les tests multi-gènes utilisés dans les laboratoires de diagnostic moléculaire en parallèle de *BRCA1* et *BRCA2*. Cependant en France comme ailleurs, l'utilité clinique de ces tests multi-gènes fait débat. En effet, les données familiales actuelles sont insuffisantes pour avoir une estimation précise du risque absolu de cancer du sein pour les porteurs. De fait, il n'y a pas de consensus quant aux recommandations de surveillance et de suivi pour ces patients et pour leurs apparentés porteurs.

On estime que 5 à 10% des cancers du sein sont d'origines familiales. Grâce aux études de liaison, d'association et aux études pangénomiques on sait aujourd'hui expliquer environ 50% du risque génétique (Figure 13). L'héritabilité manquante pourrait être expliquée par l'identification de nouveaux SNPs associés avec la maladie ou de nouveaux gènes de prédisposition de risque intermédiaires mis en évidence par le séquençage de gènes candidats ou le séquençage d'exome dans de larges études.

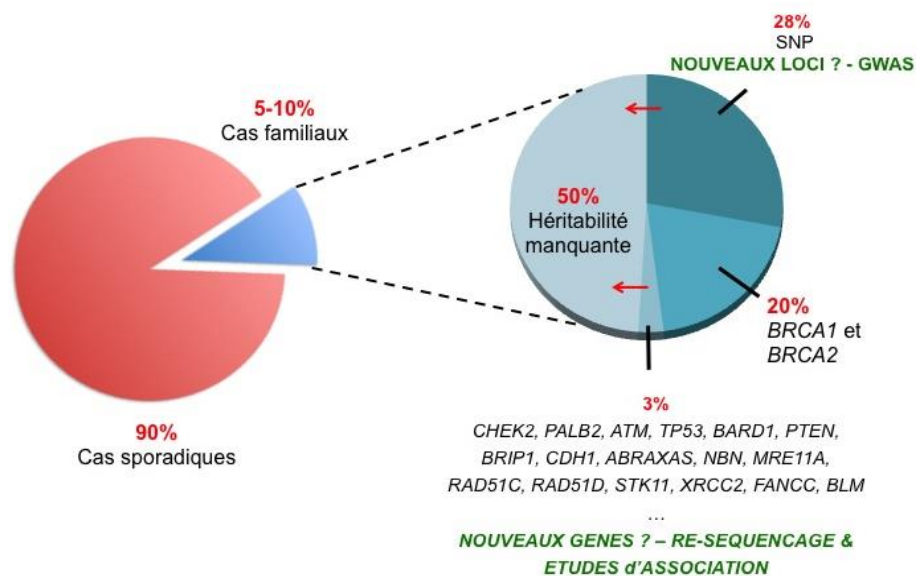


Figure 13 : Proportion des cas familiaux de cancer du sein et gènes/loci mis en évidence dans les études familiales et dans les études de population .

La découverte de nouveaux gènes/loci permettrait d'expliquer une part de l'héritabilité manquante.

4. La problématique ATM

4.1 ATM : un gène, une protéine, des fonctions

4.1.1 La protéine ATM

Le gène *ATM* (*Ataxia-Telangiectasia Mutated*) est localisé au locus 11q22 et mesure 150kb. Soixante-trois exons le composent dont 62 codant pour une protéine¹⁶⁰ longue de 3056 AA et appartenant à la famille des *Phosphatidyl inositol 3' kinase-related kinases* (PIKKs) (Figure 14). ATM possède un domaine FATC, commun aux extrémités C-terminales de toutes les protéines de la famille PIKK. Par ailleurs, quatre autres domaines fonctionnels se distinguent : le domaine NLS assurant le transport de la protéine au noyau, le domaine à leucine zipper, le domaine FAT contenant la sérine 1981 (site d'auto-phosphorylation d'ATM) et le domaine PI3-K.

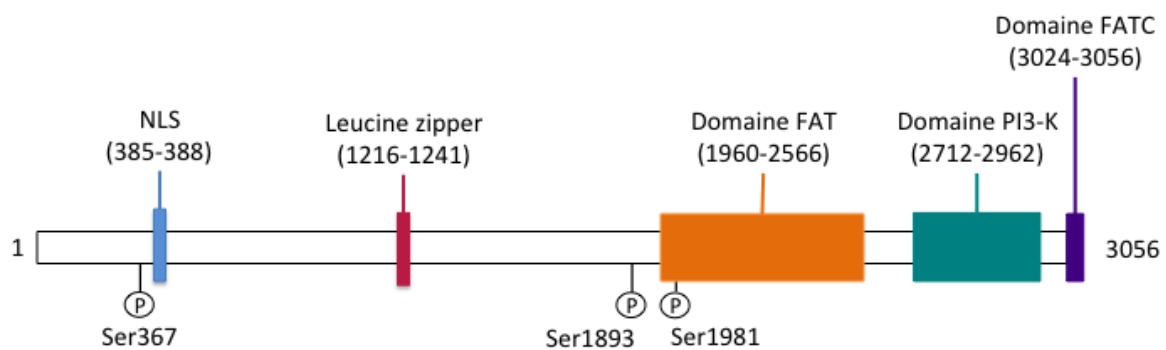


Figure 14 : Domaines fonctionnels d'ATM.

4.1.2 Les fonctions d'ATM

La protéine ATM joue un rôle majeur dans le maintien de l'intégrité du génome par son implication en tant que « transmetteur » dans le contrôle du cycle cellulaire et dans la détection des cassures doubles-brins de l'ADN (« DSB » pour *Double-Strand Break*)¹⁶¹.

Lorsque la cellule est soumise à des radiations ionisantes, un stress oxydatif ou encore des agents chimiques, des cassures DSB peuvent survenir et impliquer une réparation par recombinaison homologue (HR) ou par jonction des brins non-homologues (NHEJ, pour *Non-Homologous End-Joining*). Lorsque la cellule n'est pas exposée à ces différents stress, ATM est présente dans le noyau sous forme de dimères inactifs qui vont se dissocier par autophosphorylation d'ATM sur la sérine 1981 suite à l'induction d'un DSB¹⁶². Le recrutement d'ATM aux sites de dommages de l'ADN est conduit par le complexe MRN composé de MRE11, RAD50 et NBN. La protéine MDC1 se lie alors à ATM et stimule l'activité de phosphorylation d'ATM sur les histones H2AX qui vont en retour recruter d'avantage de protéines MDC1 au niveau des cassures DSB ainsi qu'un relâchement de la chromatine¹⁶³. Parallèlement, ATM va permettre le recrutement de la protéine BRCA1 de façon

directe ou indirecte en phosphorylant les protéines CHEK1 et CHEK2¹⁶⁴. D'autres protéines telles que p53 sont recrutées aux sites de cassures et vont entraîner l'activation des points de contrôle du cycle cellulaire, un arrêt de celui-ci et enfin la réparation des dommages à l'ADN ou l'entrée de la cellule en phase d'apoptose¹⁶⁵.

Plus de 1 000 protéines ont été identifiées comme substrats potentiels d'ATM ; ces protéines sont impliquées dans le contrôle de l'expression des gènes, dans l'organisation de la chromatine, dans la mort cellulaire programmée ou encore dans la régulation de l'élongation des télomères¹⁶⁶. Nous reviendrons sur ce dernier mécanisme un peu plus loin dans ce mémoire.

4.2 L'ataxie-télangiectasie

Des mutations bialléliques inactivatrices d'ATM sont responsables de l'ataxie-télangiectasie (A-T)¹⁶⁷, un syndrome rare touchant 1 enfant sur 100 000. Également appelé syndrome de Louis-Bar, du nom du médecin qui a décrit pour la première fois le phénotype de cette maladie, la transmission de l'A-T suit un mode autosomique récessif. La maladie se caractérise par une immunodéficience primitive touchant principalement les lymphocytes T et les anticorps¹⁶⁸, une dégénérescence neuronale progressive affectant le cervelet et entraînant un déséquilibre de la marche (ataxie)¹⁶⁹, la présence de petits vaisseaux sanguins (télangiectasies)¹⁷⁰ au niveau de la conjonctive et à la surface de la peau, et un risque accru de cancer¹⁷¹, notamment de lymphome. Les premiers symptômes de l'A-T apparaissent généralement vers 2-3 ans avec l'apprentissage de la marche chez l'enfant. Le diagnostic de l'A-T est confirmé par un dosage de l'alpha-fœto protéine supérieur à 10 ng/ml et par un déficit en immunoglobuline A et E (IgA et IgE). De plus, au niveau cellulaire, la cytologie des lymphocytes des patients A-T est caractérisée par des cassures chromosomiques principalement au niveau des chromosomes 7 et/ou 14.

La majorité des enfants atteints d'A-T sont hétérozygotes composites, c'est à dire que les mutations héritées de la mère et du père sont différentes^{172, 173}. La majorité des mutations mises en évidence chez ces enfants sont des mutations « perte de fonction » ou « tronquantes » mais une petite proportion des mutations causales sont des mutations faux-sens¹⁷⁴. Li et al. ont montré à partir des données moléculaires et cliniques de 134 enfants atteints d'A-T que les porteurs de deux mutations faux-sens avaient un phénotype atténué, avec par exemple une espérance de vie plus longue et un poids plus élevé que les porteurs de deux mutations tronquantes¹⁷³.

Par ailleurs, en raison du rôle d'ATM dans la réparation des cassures DSB de l'ADN, l'absence d'expression de la protéine dans les cellules des enfants A-T entraîne une forte instabilité génomique et une sensibilité accrue aux radiations ionisantes et aux radiations médicales¹⁷⁵.

4.3 Quels risques de cancer pour les porteurs hétérozygotes d'une mutation d'ATM ?

Dès les années 80, Swift et coll. mettent en évidence sur un échantillon de 110 familles A-T une incidence de cancers particulièrement élevée dans ces familles. Le risque relatif (RR) de cancer pour les hommes porteurs hétérozygotes d'une mutation d'ATM (HetAT) est estimé à 2,3 ($p=0,014$) et à 3,1 ($p=0,004$) pour les femmes HetAT. Dans ces familles, le RR pour le cancer du sein s'élève à 6,8 pour les femmes HetAT ($p=0,006$)¹⁷⁶. Quelques années plus tard, plusieurs études menées dans des familles A-T françaises, anglaises, américaines et norvégiennes confirment ces premiers résultats et montrent que le risque de cancer du sein des femmes HetAT est 3 à 6 fois celui de non porteuses de mutations (non-HetAT)¹⁷⁷⁻¹⁸¹ selon la population étudiée.

D'autre part, l'étude française rétrospective sur les familles A-T coordonnée par Dominique Stoppa-Lyonnet et Nadine Andrieu afin d'évaluer les risques de cancer dans ses familles a également montré que le risque de cancer du sein varie en fonction de la localisation de la mutation familiale. Les mutations localisées dans le site d'interaction d'ATM et de BRCA1 et de p53 (200 premiers acides aminés), dans le site de fixation de la B-adaptine et dans le domaine FAT sont associées à un risque de cancer du sein plus élevé que des mutations localisées ailleurs sur la protéine¹⁸².

Comme nous l'avons vu précédemment, des études de types cas-témoins menées en population générale, c'est à dire en dehors d'un contexte familial de prédisposition à l'A-T se sont également intéressées au rôle d'ATM dans la prédisposition au cancer du sein¹⁸³⁻¹⁸⁵.

4.4 Quelles implications en oncogénétique ?

D'après les études de population mentionnées plus haut, on estime que 0,5% à 1% de la population serait porteuse d'un variant délétère d'ATM ; ces variants seraient au moins trois fois plus fréquents chez les sujets atteints d'un cancer du sein d'origine familiale. Pour cette raison, comme nous l'avons vu plus haut, ATM fait partie des gènes séquencés en parallèle de BRCA1, BRCA2 et PALB2, chez les sujets issus de familles prédisposées au cancer du sein et/ou de l'ovaire, dans les laboratoires de diagnostic moléculaire. Bien qu'il n'existe pas de consensus sur les traitements et le suivi des patients, aux Etats-Unis, le NCCN recommande un suivi annuel reposant sur une mammographie ou une IRM à partir de 40 ans et une mastectomie prophylactique peut également être envisagée selon l'histoire familiale. En France, cette surveillance est appliquée chez les participantes à l'étude de suivi sur les femmes apparentées à un enfant atteint d'A-T (étude CoF-AT détaillée dans la partie II) mais aucune recommandations n'est prescrites pour les femmes porteuses d'un variant délétère d'ATM

identifiées hors contexte A-T.

De plus, aucun traitement spécifique n'est actuellement proposé aux patients HetAT en raison du manque de connaissance sur les caractéristiques histologiques et moléculaires des tumeurs du sein qu'ils développent.

5. Objectifs de la thèse

Le premier objectif de la thèse a consisté à rechercher des facteurs génétiques constitutionnels pouvant expliquer une partie de la variabilité du risque de développer un cancer chez les porteurs hétérozygotes d'une mutation dans le gène *ATM* (HetAT). Pour ce faire j'ai comparé les profils génétiques de personnes HetAT ayant développé un cancer du sein à celui de personnes indemnes. J'ai orienté mes analyses sur *i)* la recherche de variants rares potentiellement délétères et situés dans d'autres gènes de prédisposition au cancer du sein, *ii)* le rôle de la longueur des télomères qui est un marqueur du vieillissement cellulaire et un facteur de risque pour certains cancers et *iii)* sur l'implication de SNPs fréquents dans la population et ayant été associés au risque de cancer du sein chez les cas dits « sporadiques » ou modifiant le risque de cancer du sein chez les porteuses d'une mutation constitutionnelle dans le gène *BRCA1* ou *BRCA2*.

Dans un deuxième volet, je me suis attachée à décrire les caractéristiques histologiques et moléculaires des tumeurs du sein développées par les patients ayant hérité une ou deux copies mutées d'*ATM* afin d'identifier des biomarqueurs permettant de discriminer les tumeurs du sein des sujets HetAT des autres tumeurs, et donc d'évaluer le potentiel clinique de détection d'une telle signature.

**Partie II : Facteurs génétiques associés
au risque de cancer dans les familles**
ATM

1. Contexte de l'étude

La pénétrance incomplète des variants délétères dans le gène *ATM* suggère que d'autres facteurs génétiques constitutionnels augmenteraient ou diminueraient le risque de cancer chez les individus porteurs hétérozygotes d'une mutation dans ce gène dans les familles A-T. Dans la première partie de la thèse j'ai analysé trois types de facteurs génétiques constitutionnels susceptibles de moduler le risque de cancer du sein :

- la présence d'un variant rare délétère dans un autre gène connu de prédisposition au cancer du sein,
- la longueur des télomères,
- les variants fréquents dans la population (SNP) rapportés comme étant associés à la longueur des télomères ou au risque de cancer du sein dans d'autres populations.

2. Population étudiée

Pour réaliser ces travaux, je me suis appuyée sur deux ressources disponibles à l'Institut Curie : l'étude rétrospective sur les familles apparentées à un enfant atteint d'ataxie-télangiectasie (A-T) et l'étude prospective sur les femmes apparentées à un enfant atteint d'A-T.

2.1 Etude rétrospective sur les familles A-T

Réalisée entre juin 1994 et février 1997 à l'Institut Gustave Roussy et à l'Institut Curie, l'étude rétrospective sur les familles A-T (Retro-AT) est la première étude française portant sur les familles d'enfants atteints d'ataxie-télangiectasie ¹⁷⁹. Cette étude, initiée par le Dr Nadine Andrieu, le Pr Dominique Stoppa-Lyonnet et le Dr Nicolas Janin avait pour but de recenser les familles françaises A-T et d'étudier le risque de cancer pour les apparentés d'un enfant atteint d'A-T ^{181, 182, 186}.

Les enfants atteints d'A-T ont été identifiés par les cytogénéticiens et les pédiatres qui assuraient leur suivi médical. Tous les enfants atteints d'A-T vivant en France ainsi que tous leurs apparentés âgés de plus de 18 ans étaient éligibles pour l'étude (parents, grands-parents, oncles, tantes, cousins, cousines...).

Trente-quatre familles A-T ont ainsi été ainsi recrutées, ce qui représente 1 429 sujets dont 44 enfants atteints d'A-T, 711 femmes et 674 hommes. Un échantillon sanguin ou salivaire a été prélevé à l'inclusion chez les enfants et chez 402 participants afin de déterminer leur statut mutationnel pour *ATM*. Les allèles mutés ont été identifiés chez les enfants par génotypage de 8 marqueurs microsatellites repartis entre les loci des gènes *MRE11* et *ATM* : D11S1817, D11S1819, D11S1778, D11S2180, D11S1294, D11S2180, D11S2178 et D11S2179. La ségrégation des allèles mutés dans les branches maternelles et paternelles a ensuite été déterminée par le génotypage des participants

pour les marqueurs microsatellites associés à l'A-T.

2.2 Etude prospective sur les familles A-T : CoF-AT

2.2.1 La cohorte CoF-AT

L'étude CoF-AT pour **C**ohorte prospective de **F**emmes apparentées à un enfant atteint d'**A**taxie-**T**élangiectasie a été initiée en 2003 par le Dr Nadine Andrieu et le Pr Dominique Stoppa-Lyonnet. Ayant déjà confirmé par l'étude Retro-AT l'hypothèse d'une prévalence plus forte de cancer du sein dans les familles françaises A-T ¹⁸⁶, cette étude vise à mieux comprendre le rôle du gène *ATM* dans l'étiologie du cancer du sein. Les objectifs de l'étude sont multiples :

- Proposer aux femmes un suivi permettant un dépistage précoce de cancer du sein.
- Estimer plus précisément le risque de cancer associé au gène *ATM*, en particulier le risque de cancer du sein.
- Etudier l'histoire naturelle du cancer du sein chez les femmes porteuses d'une mutation à l'état hétérozygote dans le gène *ATM*.

L'étude CoF-AT est coordonnée par la Plateforme d'Investigation en Epidémiologie Génétique (PIGE) intégrée au sein de l'équipe Epidémiologie Génétique des Cancers de l'U900. L'équipe coordonnatrice est composée d'Eve Cavaciuti et de Dorothée Le Gal qui réalisent l'inclusion et le suivi des participantes, collectent et codent dans la base les données familiales, cliniques et épidémiologiques et contactent les médecins pour valider les informations médicales fournies par les participantes.

2.2.2 Identification des familles A-T

Parmi les 34 familles identifiées via l'étude Retro-AT, 27 d'entre elles ont accepté de participer à l'étude CoF-AT. L'identification des nouvelles familles A-T se fait grâce aux pédiatres, neuro-pédiatres et à l'association pour la recherche sur l'A-T (APRAT : <http://www.aprat.fr>). Le diagnostic phénotypique posé par les médecins est validé par un diagnostic moléculaire réalisé dans le service de génétique de l'Institut Curie (IC), centre de référence en France pour les tests génétiques sur le gène *ATM*. Un test ciblé est également effectué chez les parents pour déterminer l'origine paternelle et maternelle des mutations responsables de la maladie de l'enfant. Le service de génétique de l'IC informe le centre coordonnateur de l'identification d'une nouvelle famille A-T et transmet les coordonnées du médecin afin de l'informer sur l'étude qu'il pourra alors proposer à la famille.

2.2.3 L'inclusion des participantes

Toutes les femmes apparentées à un enfant atteint d'A-T (la mère, les sœurs, les grands-mères

maternelles et paternelles, les tantes maternelles et paternelles, les nièces) de plus de 18 ans et indemnes de tout cancer sont éligibles pour participer à l'étude. La première femme incluse dans l'étude, généralement la mère de l'enfant atteint d'A-T, reçoit des fiches d'information (Annexe 2) et de pré-participation (Annexe 3) à transmettre aux autres femmes de la famille éligibles pour l'étude. Les inclusions se font ensuite en cascade grâce aux femmes déjà incluses qui communiquent sur l'étude au sein de leur famille.

Les participantes sont incluses lors d'une consultation d'oncogénétique et de génétique au cours de laquelle les femmes signent un consentement de participation (Annexe 4) et élaborent avec l'oncogénéticien(ne) un arbre aussi détaillé que possible sur les antécédents familiaux dans les branches maternelles et paternelles. Pour chaque membre de la famille répertorié sur l'arbre généalogique la date de naissance, l'âge (ou la date de décès) et le sexe sont demandés. Pour les antécédents médicaux, notamment de cancer, une fiche d'autorisation d'accès au dossier médical est demandée afin de confirmer la date de diagnostic et la localisation du cancer. Lorsque j'ai initié ce projet de thèse en novembre 2013, 415 femmes participaient à l'étude ; aujourd'hui (juillet 2017) le nombre de femmes suivies s'élève à 509.

2.2.4 Statut mutationnel d'*ATM* des participantes

Un prélèvement sanguin est réalisé lors de l'inclusion afin de déterminer le statut mutationnel d'*ATM* des participantes selon leur appartenance à la branche maternelle ou paternelle de l'enfant A-T. Le test génétique est effectué dans le service de génétique de l'IC où les échantillons de sang et d'ADNs sont centralisés.

Parmi les 415 participantes étudiées dans le cadre de la thèse, 213 femmes sont porteuses de la mutation familiale d'*ATM* (ces femmes seront appelées par la suite « HetAT ») et 202 femmes ne portent pas cette mutation (ces femmes seront appelées « non-HetAT »).

2.2.5 Suivi des participantes

À l'inclusion, les participantes remplissent un premier questionnaire épidémiologique (EPI1) (Annexe 5) compilant des questions sur leurs antécédents médicaux et gynécologiques ainsi que sur leur mode de vie (consommation de tabac et expositions aux irradiations). Puis, tous les deux ans pendant 10 ans les participantes reçoivent un nouveau questionnaire (EPI2, EPI3, EPI4, EPI5 et EPI6) reprenant les questions du premier questionnaire auxquelles s'ajoutent de nouvelles questions sur la consommation d'alcool, la prise de Distilbène chez la mère de la participante et l'origine ethnique (Annexe 6). Les questionnaires sont revus avec les femmes lors d'une interview téléphonique par le centre coordinateur pour confirmer, et préciser si besoin, les informations fournies.

Un suivi médical est également proposé aux participantes âgées de 25 ans et plus. Ce suivi consiste

pour les femmes de 25 à 40 ans en une surveillance clinique mammaire seule, et pour les femmes de plus de 40 ans une mammographie est prescrite tous les deux ans.

2.3 Autre source d'échantillons biologiques : familles A-T recrutées lors d'une consultation de génétique à l'Institut Curie

Le service de génétique de l'institut Curie est le centre de référence pour le diagnostic moléculaire de l'A-T. Le résultat du test génétique pour l'A-T est divulgué aux parents de l'enfant atteint d'A-T lors d'une consultation de génétique. Au cours de l'entretien, le médecin leur propose de signer un consentement éclairé autorisant l'utilisation de leur prélèvement sanguin et de ses dérivés à des fins de recherche. Des sujets de 47 familles A-T ne participant ni CoF-AT ni Retro-AT ont pu ainsi être inclus dans les analyses moléculaires réalisées dans le cadre de cette thèse (11 femmes HetAT, 52 hommes HetAT ainsi qu'une femme et trois hommes non-HetAT).

2.4 Prévalence des cancers dans les familles participant à CoF-AT ou à Retro-AT

Lors de l'initiation de mon projet de thèse, 45 femmes de l'étude CoF-AT avaient développé un cancer après leur inclusion dans l'étude. Vingt huit de ces femmes avaient développé un cancer du sein et 20 d'entre elles étaient porteuses d'une mutation d'*ATM* (Tableau 1). Il y avait donc deux fois plus de femmes HetAT que de femmes non-HetAT qui avaient développé la maladie.

Par ailleurs, parmi les hommes et les femmes ayant participé uniquement à l'étude Retro-AT, 38 participants HetAT (12 femmes et 26 hommes) et 48 participants non-HetAT (32 femmes et 16 hommes) ont été atteints d'un cancer (Tableau 7). On note que le nombre de cas de cancer chez les femmes HetAT est inférieur au nombre de cas observés chez les femmes non-HetAT, ce qui pourrait s'expliquer par une volonté plus forte des femmes HetAT de participer à l'étude Retro-AT.

Grâce au suivi des familles de CoF-AT, le centre coordonnateur a également recensé les enfants et adultes A-T ayant développé un cancer. Parmi les 37 patients A-T atteints d'un cancer, trois d'entre eux ont développé un cancer du sein, dont un homme (Tableau 7).

Tableau 7 : Prévalence des cancers chez les participants aux études CoF-AT et Retro-AT.

	Cohorte CoF-AT		Etude Retro-AT		Enfants atteints d'A-T
	Non-HetAT (N=14)	HetAT (N=31)	Non-HetAT (N=32)	HetAT (N=38)	Homozygotes composites (N=37)
Femmes atteintes	N=14	N= 31	N=16	N=12	N=17
Col de l'utérus	0	2	1	1	0
Colon-rectum	1	0	0	0	0
Estomac	0	0	0	0	1
Foie	0	0	0	1	0
Leucémie	0	0	0	0	1
Leucémie lymphoïde	0	0	0	0	1
Lèvre-bouche-larynx	0	0	1	0	0
Lymphome non-Hodgkinien	1	1	1	1	8
Maladie de Hodgkin	0	0	0	0	1
Maladie immunoproliférative maligne	0	0	0	0	1
Mélanome	0	1	0	0	0
Organe digestif	0	0	1	2	0
Organes génitaux	0	1	0	0	0
Ovaire	1	0	1	0	0
Pancréas	0	1	2	0	0
Poumon	1	0	1	0	0
Rein	2	1	1	0	0
Sein	8	20	6	5	2
Thyroïde	0	0	1	0	1
Utérus	1	1	1	0	0
Vésicule biliaire	0	0	0	1	0
Site du cancer primitif inconnu	0	0	0	0	1
Hommes atteints	0	0	N=16	N=26	N=20
Colon-Rectum	0	0	2	0	0
Estomac	0	0	1	2	0
Foie	0	0	0	0	1
Leucémie	0	0	0	1	2
Leucémie lymphoïde	0	0	1	0	4
Leucémie myéloïde	0	0	1	1	0
Lèvre-bouche-larynx	0	0	2	1	0
Lymphome non-Hodgkinien	0	0	0	1	7
Maladie de Hodgkin	0	0	0	1	1
Mélanome	0	0	0	1	0
Organes digestifs	0	0	0	0	0
Pancréas	0	0	1	2	0
Poumon	0	0	0	4	0
Prostate	0	0	4	10	0
Sein	0	0	0	0	1
Testicules	0	0	0	1	0
Tissus conjonctif	0	0	0	0	2
Vessie	0	0	0	1	0
Site du cancer primitif inconnu	0	0	2	0	2

3. Recherche de variants rares délétères dans des gènes connus de prédisposition au cancer autre que le gène *ATM*

Lors de l'inclusion dans l'étude Retro-AT ou CoF-AT seul le statut mutationnel d'*ATM* des participants est déterminé. Hormis une famille A-T connue pour ségréger, en plus des mutations dans *ATM*, une mutation délétère dans le gène *BRCA1* et deux autres familles A-T ségrégant également une mutation délétère dans le gène *BRCA2*, aucune autre altération génétique constitutionnelle n'avait été recherchée pour expliquer la survenue des cancers dans les familles participantes.

Afin d'examiner la présence de potentiels variants rares délétères, localisés dans des gènes de prédisposition au cancer du sein j'ai utilisé le test multi-gènes développé par le service de génétique de l'institut Curie et utilisé depuis 2015 pour le diagnostic moléculaire des familles hautement prédisposées au cancer. Le panel de gènes utilisé par le laboratoire est composé de 55 gènes associés aux formes familiales de cancer du sein et de l'ovaire, au rétinoblastome, aux syndromes de l'anémie de Fanconi, de Lynch, de Peutz-Jeghers, de Cowden, de Bloom, de Nijmegen et de Li-Fraumeni. Le panel de gène inclus également des gènes d'intérêt pour la recherche en épidémiologie génétique. La liste complète des gènes ciblés par ce panel est détaillée dans le tableau 8.

3.1 Sélection des sujets d'étude

Parmi les 152 cas de cancer recensés dans les études CoF-AT et Retro-AT j'ai sélectionné, indépendamment du statut mutationnel pour *ATM*, les 34 femmes et les deux patients A-T ayant développé un cancer du sein et pour lesquels un échantillon d'ADN était disponible au laboratoire (Tableau 9). J'ai également analysé l'ADN constitutionnel de 15 femmes et cinq patientes A-T ayant développé un cancer à une autre localisation que le sein. Après dosage des échantillons, j'ai dû exclure de l'analyse huit patients pour lesquels la quantité d'ADN disponible s'est avérée insuffisante pour le séquençage (Tableau 9).

Finalement, les 55 gènes ont été séquencés chez 48 participants dont 30 femmes HetAT (dont 21 avaient développée un cancer du sein), 12 femmes non-HetAT (dont cinq avaient développé un cancer du sein) et six patients A-T (dont un avait développé un cancer du sein) (Tableau 9).

Tableau 8 : Liste des 55 gènes séquencés et maladies héréditaires associées

	Gènes	Maladie associée	Autres cancers associés
Gènes de prédisposition de forte pénétrance	BRCA1		Ovaire
	BRCA2		Ovaire
	PALB2		Ovaire
	CDKN2A	Mélanome	Sein
Gènes de prédisposition au cancer du sein de pénétrance modérée	ATM	Ataxie-télangiectasie	
	CHEK2		
	RAD50		
	RAD51C		Sein/ovaire
	RAD51D		Ovaire
	RINT1		Sein/Lynch
	XRCC2		
Gènes de prédisposition au cancer du colon	APC		
	MLH1	Syndrome de Lynch	
	MSH2	Syndrome de Lynch	
	MSH6	Syndrome de Lynch	Endomètre
	EPCAM	Syndrome de Lynch	Endomètre
	PMS2	Syndrome de Lynch	
Gènes associés à des syndromes	MUTYH		
	CDH1		Sein (lobulaire)
	BLM	Syndrome de Bloom	Sein
	MRE11A	Ataxie-télangiectasie « like »	Sein
	NBN	Syndrome de Nijmegen	Sein
	PTEN	Syndrome de Cowden	Sein, Thyroïde, Utérus, Ovaires
	STK11	Syndrome de Peutz-Jeghers	Sein, Colorectal
	TP53	Syndrome de Li-Fraumeni	Sein, Leucémie, Cerveau, Sarcomes
	FANCA	Anémie de Fanconi	
	FANCB	Anémie de Fanconi	
	FANCC	Anémie de Fanconi	
	FANCD2	Anémie de Fanconi	
	FANCE	Anémie de Fanconi	
	FANCF	Anémie de Fanconi	
	FANCG/XRCC9	Anémie de Fanconi	
	FANCI	Anémie de Fanconi	
	FANCI/BRIP1	Anémie de Fanconi	
	FANCL	Anémie de Fanconi	
	FANCM	Anémie de Fanconi	
	FANCN/PALB2	Anémie de Fanconi	
FANCP/SLX4	Anémie de Fanconi		
FANCC/ERCC4	Anémie de Fanconi		
Gènes de prédisposition au rétinoblastome	RB1		
Gènes d'intérêt pour la recherche	MED4		
	BARD1		
	BAP1		
	BMPR1A		
	DICER1	Dysplasie Pleuropulmonaire	
	FAN1		
	HELQ		
	MDM2		Variants associés à la survie des patientes affectées par un cancer du sein
	MET		Sein, gastro-intestinal, mélanome, poumon
	POLD1		Colorectal
	POLE		Colorectal
	RAD51B		
	SLC4A7		
	SMAD4	Polypose juvénile	Colorectal
	SMARCB1		Rhabdoïde
	WT1	Tumeur de Wilms	
	XRCC3		

Tableau 9 : Liste des participants ayant développé un cancer chez lesquels les 55 gènes de prédisposition au cancer ont été analysés.

Patient	Etude	Statut ATM	Sexe	Cancer	Age au diagnostic	Séquencé
AT001006	CoF-AT	Hétérozygote	F	Sein	62	oui
AT001007	CoF-AT	Hétérozygote	F	Sein	40	oui
AT021005	CoF-AT	Hétérozygote	F	Sein	60	oui
AT022017	CoF-AT	Hétérozygote	F	Sein	53	oui
AT023006	CoF-AT	Hétérozygote	F	Sein	45	oui
AT035001	CoF-AT	Hétérozygote	F	Sein	55	oui
AT047004	CoF-AT	Hétérozygote	F	Sein	47	oui
AT082001	CoF-AT	Hétérozygote	F	Sein	65	oui
AT087001	CoF-AT	Hétérozygote	F	Sein	66	oui
AT088046	CoF-AT	Hétérozygote	F	Sein	50	oui
AT105022	CoF-AT	Hétérozygote	F	Sein	51	oui
AT107020	CoF-AT	Hétérozygote	F	Sein	74	oui
AT107021	CoF-AT	Hétérozygote	F	Sein	30	oui
AT120001	CoF-AT	Hétérozygote	F	Sein	49	non
AT138021	CoF-AT	Hétérozygote	F	Sein	52	oui
AT149008	CoF-AT	Hétérozygote	F	Sein	45	oui
AT161008	CoF-AT	Hétérozygote	F	Sein	74	non
AT215996	CoF-AT	Hétérozygote	F	Sein	40	oui
AT520001	CoF-AT	Hétérozygote	F	Sein	62	oui
AT541006	CoF-AT	Hétérozygote	F	Sein	36	oui
AT001045	CoF-AT	Non porteur	F	Sein	42	non
AT063051	CoF-AT	Non porteur	F	Sein	42	oui
AT087006	CoF-AT	Non porteur	F	Sein	50	oui
AT098041	CoF-AT	Non porteur	F	Sein	58	non
AT108041	CoF-AT	Non porteur	F	Sein	50	oui
AT138022	CoF-AT	Non porteur	F	Sein	36	oui
AT204015	CoF-AT	Non porteur	F	Sein	48	oui
AT001057	CoF-AT	Hétérozygote	F	Organes génitaux	18	oui
AT004007	CoF-AT	Hétérozygote	F	Thyroïde	47	oui
AT004009	CoF-AT	Hétérozygote	F	Thyroïde	39	oui
AT009038	CoF-AT	Hétérozygote	F	Col de l'utérus	33	oui
AT023026	CoF-AT	Hétérozygote	F	Rein	73	oui
AT035006	CoF-AT	Hétérozygote	F	Lymphome non Hodgkinien	78	oui
AT047002	CoF-AT	Hétérozygote	F	Pancréas	73	oui
AT090001	CoF-AT	Hétérozygote	F	Mélanome	52	non
AT127009	CoF-AT	Hétérozygote	F	Lymphome non Hodgkinien	43	oui
AT206040	CoF-AT	Hétérozygote	F	Colon Rectum	52	oui
AT141010	CoF-AT	Non porteur	F	Lymphome non Hodgkinien	47	oui
AT162003	CoF-AT	Non porteur	F	Rein	70	oui
AT242004	CoF-AT	Non porteur	F	Ovaires	33	oui

AT273025	CoF-AT	Non porteur	F	Rein	74	oui
AT404002	CoF-AT	Non porteur	F	Poumon	49	oui
AT063003	Retro-AT	Hétérozygote	F	Sein	35	oui
AT088042	Retro-AT	Hétérozygote	F	Sein	77	oui
AT114011	Retro-AT	Hétérozygote	F	Sein	58	oui
AT001011	Retro-AT	Non porteur	F	Sein	96	non
AT001012	Retro-AT	Non porteur	F	Sein	44	oui
AT103032	Retro-AT	Non porteur	F	Sein	67	non
AT096024	Retro-AT	Non porteur	F	Utérus/Sein	64/66	oui
AT021001	Patient AT	Hétérozygote composite	H	Sein	28	oui
AT137001	Patient AT	Hétérozygote composite	F	Sein	42	non
AT105001	Patient AT	Hétérozygote composite	F	Lymphome non Hodgkinien	13	oui
AT186002	Patient AT	Hétérozygote composite	F	Estomac	32	oui
AT193007	Patient AT	Hétérozygote composite	F	Thyroïde	22	oui
AT211003	Patient AT	Hétérozygote composite	F	Maladie immunoproliférative maligne	25	oui
AT269004	Patient AT	Hétérozygote composite	F	Lymphome non Hodgkinien	10	oui

Huits patients n'ont pas été séquencés faute de quantité d'ADN suffisante pour réaliser le séquençage à haut débit.

3.2 Préparation des banques d'ADN et séquençage.

Les exons des 55 gènes ont été capturés par enrichissement des régions codantes en utilisant le protocole SureSelect^{QXT} Target Enrichment system (Agilent Technologies) (Figure 14) qui permet de cibler des régions d'intérêt. Les échantillons d'ADN sont préparés en même temps que deux échantillons témoins présentant respectivement une grande délétion des exons 2 à 6 dans le gène *BRCA1* et une mutation délétère dans le gène *RB1* (c.52insCCCTCCTGAG ; p.Glu19Leufs15X). Avant la préparation des banques, les échantillons d'ADN génomique (ADNg) sont dosés à l'aide du kit Qubit dsDNA BR permettant de doser spécifiquement l'ADN double-brin. Les échantillons sont dilués à 25ng/μL dans un volume final de 10μL.

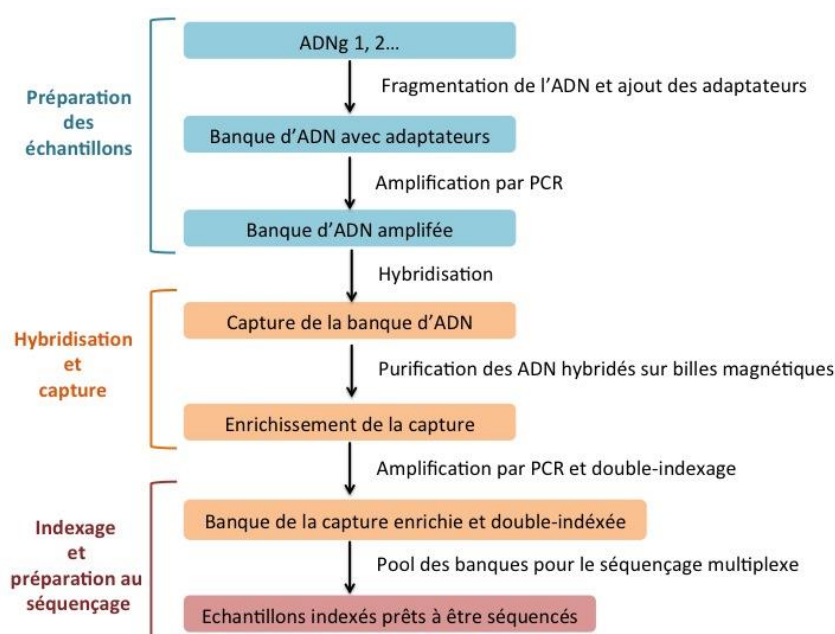


Figure 15 : Processus de préparation des échantillons d'ADN avec le kit *SureSelect Target Enrichment* d'Agilent.

3.2.1 Préparation des échantillons

Cette première étape vise à amplifier l'ADNg pour créer une banque d'ADN qui facilitera la capture des amplicons ciblés. Cette étape se décompose en 3 parties : la fragmentation de l'ADN, l'ajout d'adaptateurs et l'amplification par PCR (Figure 13).

Vingt-cinq nanogrammes d'ADNg de chaque échantillon sont digérés par une réaction enzymatique générant des fragments de 300pb. Parallèlement à la digestion, deux adaptateurs A (AGATCGGAAGAGCACACGTCTCTGAACTCCAGT) et B (AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT) de 33pb sont ajoutés aux extrémités de chaque fragments afin de servir d'amorces pour l'amplification de la banque d'ADN par réaction en chaîne par polymérase (PCR) (NB : le protocole complet de la préparation des banques d'ADN figure en Annexe 7).

Les banques d'ADNg sont ensuite purifiées sur billes magnétiques par 2 lavages successifs à l'éthanol 70% puis éluées dans un tampon spécifique au protocole utilisé. Un premier contrôle qualité a alors été effectué par un dosage au LabChip afin de vérifier les aspects qualitatifs et quantitatifs des banques.

3.2.2 Hybridation et amplification de la capture

L'objectif de cette deuxième étape est de capturer la région codante des 55 gènes d'intérêt puis de les amplifier par PCR (Figure 14). Cette étape est réalisée à partir de 250ng des banques d'ADN amplifiées qui sont hybridées à des appâts d'ARN mesurant environ 120 nucléotides et ciblant spécifiquement les régions d'intérêt. Les ARNs étant biotinylés, les ADNs hybridés (ou la capture) sont purifiés par fixation de la biotine sur des billes de streptavidine activées. Les ADN hybridés purifiés sont ensuite amplifiés par PCR (Annexe 7).

3.2.2.1 Indexage et préparation au séquençage

Lors de l'amplification de la capture les échantillons sont « étiquetés » par un système de double index (petites séquences d'ADN) (indexés P7 et P5, Annexes 8) qui permet de réduire le taux d'erreur d'attribution des échantillons lors de l'analyse des données. Au terme de l'amplification, les banques indexées sont purifiées sur billes magnétiques, puis des contrôles quantitatifs et qualitatifs sont réalisés par dosage au Qubit dsDNA BR et par une analyse au LabChip pour évaluer la qualité des banques avant leur séquençage. Afin de ne pas séquencer préférentiellement certains échantillons, les banques obtenues pour chacun d'eux sont combinées en un mélange équimolaire dont l'équimolarité a été validée par PCR-quantitative (q-PCR). La banque à façon est ensuite diluée à 4nM et dénaturée chimiquement.

3.2.2.2 Séquençage de la banque

La banque à façon ainsi qu'une banque servant de contrôle interne (banque PhiX) sont diluées à 20pM avant d'être déposées sur la puce de séquençage Illumina. Les échantillons ont été séquencés sur l'appareil NextSeq500 (Illumina) en utilisant le kit *SureSelectQXT Read Primers with Illumina's Paired-End Cluster Generation* selon les indications du fabricant. La réaction de séquençage est réalisée de façon « paired-end » (séquençage du brin sens et anti-sens de l'ADN) pendant 218 cycles.

3.3 Analyses bioinformatiques

3.3.1 Traitement des données brutes

Les données de séquençage générées ne peuvent être analysées directement. Elles sont d'abord traitées par un processus standardisé mis en place par la plateforme de bioinformatique de l'U900 qui effectue également plusieurs étapes de contrôle qualité avant d'obtenir une liste exhaustive de variants (Figure 15).

La première étape consiste à éliminer les doubles-index qui se trouvent aux extrémités de chaque lecture ou « read » et de transformer les fichiers bruts (fichiers .bam) au format fastq. Ce premier traitement des données est réalisé avec le logiciel Consensus Assessment of Sequence and Variation (CASAVA v.2.15.0.4) d'Illumina. Les adaptateurs sont ensuite éliminés avec le logiciel Cutadapt (version 1.8) en appliquant les paramètres suivants : au moins 3pb de chevauchement entre la séquence connue de l'adaptateur et le read, et une longueur minimale de l'adaptateur de 10pb. Au terme de ces deux premières étapes la qualité des reads est évaluée avec le logiciel FASTQC (v.0.10.1).

Les reads sont ensuite alignés sur le génome humain de référence (hg19) grâce au logiciel Bowtie2 (v.2.1.0) par un système de point d'ancrage. Vingt-deux paires de base sont alignées sur le génome et servent de point d'ancrage autour duquel le read est progressivement aligné jusqu'à alignement complet. Pour chaque segment de séquence de 22 paires de bases (pb) correctement alignées le logiciel tolère le non-appariement d'une base. Si lors de l'alignement des insertions ou des délétions trop longues (plus de quelques nucléotides) sont introduites, le score d'alignement change et les reads sont éliminés.

Parmi les reads correctement alignés sur le génome humain seuls ceux ciblant les gènes du panel sont conservés. La suite de logiciels Samtools (V.1.1) est utilisée pour reconnaître et conserver les reads correspondants à la cible choisie. La qualité de couverture des régions ciblées est estimée avec le script Coverage Marker (BedTools V2.17.0) pour les régions couvertes à moins de 30X et avec le script Coverage Percentage pour les régions couvertes à au moins 30X.

3.3.2 Identification et annotation des variants

La dernière partie du traitement des données consiste à identifier et à annoter les variants. L'appel des variants, ou « variant calling », est réalisé avec le logiciel VarScan2 (v2.3.6) en appliquant les paramètres suivants :

- Une profondeur minimum de 10X pour annoter le variant.
- Au moins 2 reads portant le variant.

- Un ratio allélique d'un minimum de 15% est appliqué pour la majorité des gènes. Cependant un ratio allélique de 5% est accepté pour les variants des gènes *RB1*, *DICER1*, *BRCA1*, *BRCA2* et *APC* afin de détecter du mosaïcisme germinale.

Cette stratégie permet d'identifier les variants nucléotidiques (*Single Nucleotide Variants ou SNV*) et les insertions délétions (Indels) de moins de 15pb.

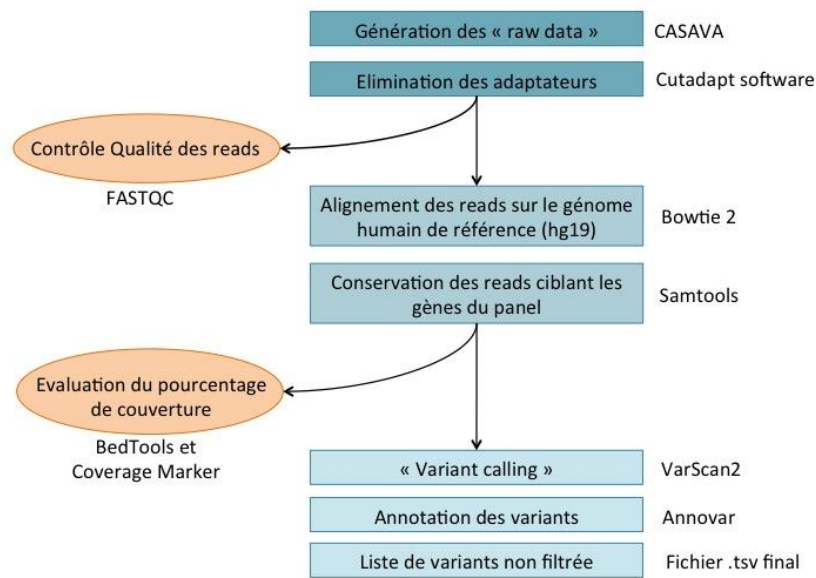


Figure 16 : Pipeline bioinformatique utilisé pour lister les variants identifiés dans les échantillons séquençés avec le panel de gènes.

Sont indiqués à gauche les logiciels utilisés pour effectuer chacune des étapes de traitement des données de séquençage (cadres bleus) ainsi que les étapes de contrôles qualité (cercles oranges).

Les variants ainsi identifiés sont annotés avec le logiciel Annovar à partir des bases de données publiques répertoriant les variants mis en évidence dans le cadre de projets spécifiques :

- dbSNP138 (NCBI) : base de données incrémentée par les chercheurs et les centres de séquençage.
- ESP6500 (National Heart, Lung and Blood Institute) : projet de séquençage de 6 500 exomes d'individus américains pour identifier de nouveaux gènes impliqués dans les maladies du cœur, des poumons et du sang.
- 1000Genomes (version 2014) : projet de séquençage du génome de 1000 individus à fin d'identifier des variants ayant une fréquence allélique d'au moins 1% dans la population générale et dans certains groupes de population. Vingt-six populations d'origines caucasiennes, afro-américaines, d'Asie de l'est et du sud sont représentées dans ce projet.

L'impact des variants exoniques sur la protéine est également évalué avec différents outils de prédiction *in silico* : SIFT (Sort Intolerant From Tolerant amino acid) ¹⁸⁷, PolyPhen2 ¹⁸⁸ et PhyloP ¹⁸⁹ (Annexe 7). Ces trois algorithmes reposent sur le fait que plus un acide aminé est conservé au cours de l'évolution plus une modification de cet acide aminé aura de conséquences sur la fonction ou la conformation de la protéine. SIFT classe les substitutions faux-sens en deux groupes : variants « tolérés » ou « délétères » d'un point de vue évolutif, PolyPhen les classe en 3 groupes : variants « bénin », « possiblement délétères » ou « probablement délétères », et PhyloP détermine si la position de l'acide aminé est conservé (score négatif) ou non (score positif) dans la phylogénie.

Le « variant calling » des insertions/délétions de taille intermédiaire (15-20pb) est réalisé séparément avec logiciel Pindel (v0.2.5a3) sur des fenêtres de 500pb, en acceptant un score de qualité d'alignement (MAPQ) d'au moins 20 (c'est-à-dire en tolérant une probabilité d'erreur de 1% selon le codage PHRED). Seuls les ratios alléliques supérieurs à 5% sont rapportés. Les réarrangements de grande taille (RGT) sont identifiés avec le logiciel DESeq en estimant le nombre de reads par exon, ou par fenêtre de 300pb pour les exons de grande taille.

L'ensemble des variants annotés est résumé dans un fichier .csv qui est ensuite importé dans la base de données développée pour le service de génétique constitutionnelle de l'Institut Curie. Celle-ci permet de lister les variants mis en évidence dans chaque échantillon.

3.3.3 Filtrage des variants

Pour déterminer quels variants pouvaient être d'intérêt dans l'implication dans le risque de cancer chez les femmes HetAT, j'ai utilisé l'algorithme de filtrage des variants mis en place par l'équipe du service de génétique constitutionnelle (Figure 16) basé sur :

- 1) La fréquence du variant dans les populations de 1000 Genomes ou dans la base de données du laboratoire de génétique. J'ai sélectionné les variants ayant une FAM < 0.5% des populations de 1000 Genomes.
- 2) Le taux de mutation de la variation, c'est à dire l'occurrence de l'altération (nombre de fois où l'altération a été observée dans la base de donnée du laboratoire de génétique) et la récurrence (pourcentage de reads portant le variant) :
 - Si l'occurrence est inférieure à 3 porteurs et si le variant est présent sur au moins 40% des reads, le variant est alors considéré comme un variant « isolé fort », suggérant qu'il s'agit d'une vraie variation de séquence potentiellement délétère. Si le variant est présents dans moins de 40% des reads il est considéré comme « isolé faible ».

- Si l'occurrence est supérieure ou égale 3, le variant est considéré comme « récurrent fort » s'il est présent dans au moins 40% des reads et comme « récurrent faible » s'il est présent dans 15 à 40% des reads.
- 3) La qualité de séquençage, reposant sur l'équilibre entre les reads forwards (F) et reverses (R) est estimée pour les variants « récurrents faibles ». Un variant est considéré comme « récurrent faible équilibré » si l'une des quatre conditions suivantes est respectée :
- La fréquence de reads F portant l'altération (F_{alt}) est comprise entre 0,1 et 0,9.
 - La fréquence de reads F_{alt} et la fréquence de reads portant l'allèle normal (F_{ref}) sont inférieures à 0.1.
 - La fréquence de reads F_{alt} et la fréquence de reads portant l'allèle normal (F_{ref}) sont supérieures à 0.9
 - Le ratio $(F_{ref}/R_{ref}) / (F_{alt}/R_{alt})$ est compris entre 0.32 et 2.96. Ces bornes ont été définies par le laboratoire de génétique en utilisant la loi de Student pour évaluer la distribution de ce ratio pour des variants connus utilisés comme contrôles. L'intervalle [0.32-2.96] a été déterminé avec ces « vrais » variants avec un risque d'erreur de 1/10 000.

Si l'une de ces conditions n'est pas respectée alors l'altération est considérée comme «récurrente faible déséquilibrée » et donc probablement non délétère.

Dans le cadre de la question posée ici (c'est à dire la recherche de variants potentiellement délétères dans un gène connu de prédisposition au cancer autre qu'*ATM* et pouvant expliquer la survenue des cancers dans les familles de CoF-AT), une attention particulière a été portée aux variants rares potentiellement délétères, définis comme « variant isolé fort » et « variant récurrent fort ».

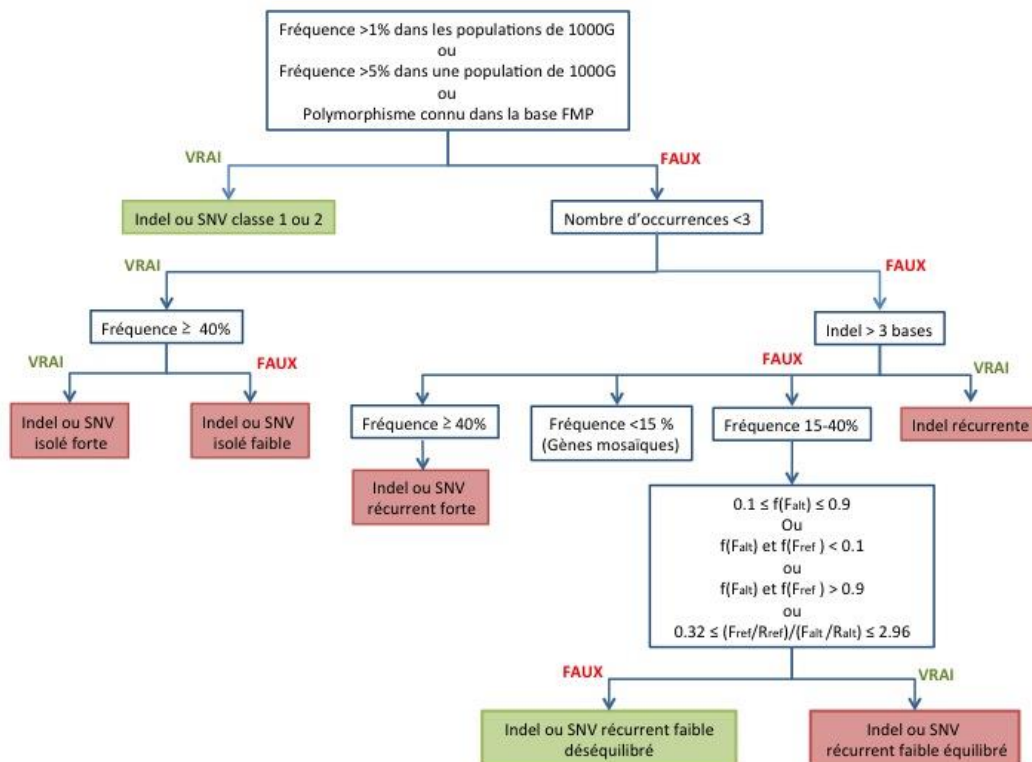


Figure 17 : Filtrés appliqués dans FileMakerPro pour classer les altérations identifiées lors du séquençage d'un panel de gènes de prédisposition au cancer.

Indel : insertion délétion ; SNV : Single Nucléotide Variant ; F : forward ; R : reverse ; $f(F_{alt})$: fréquence de reads forward portant l'altération ; $f(F_{ref})$ fréquence de reads forward portant l'allèle de référence. En rouge : variants rares à confirmer par séquençage de Sanger. En vert : variants considérés comme non délétères.

3.3.4 Classification des variants

Pour évaluer l'impact biologique des variants identifiés je me suis appuyée sur différents outils de prédiction *in silico* et sur des bases de données publiques faisant le lien entre un variant et son impact clinique possible ou avéré.

3.3.4.1 Prédiction sur les substitutions faux-sens

L'impact des substitutions faux-sens localisées dans les parties exoniques des gènes est déterminé lors de l'annotation des variants avec SIFT, PolyPhen et PhyloP (partie I.A.3.ii). Cependant, pour certains gènes comme *BRCA1*, *BRCA2*, *PALB2*, *ATM*, ou encore les gènes MMR (MisPatch Repair) (*MSH2*, *MLH1*, *MSH6* et *PMS2*) des alignements protéiques sont disponibles sur le site de l'outil de prédiction Align-GVGD (http://agvgd.hci.utah.edu/agvgd_input.php)¹⁹⁰. Cet algorithme repose sur la conservation de la protéine dans la phylogénie et prend également en compte la composition de l'acide aminé qui est substitué, son volume et sa polarité. L'avantage d'Align-GVGD par rapport à d'autres outils disponibles est qu'il fait appel à des alignements protéiques multiples qui ont été établis et corrigés manuellement afin d'obtenir un degré d'informativité (utilisation d'espèces

suffisamment éloignées pour ne pas introduire de biais dans les prédictions et surestimer l'effet délétère des variants). Ces alignements corrigés sont disponibles sur le site d'Align-GVGD (http://agvgd.hci.utah.edu/agvgd_input.php).

3.4 Résultats de l'analyse sur l'identification de nouveaux variants rares chez les individus séquencés

Le séquençage du panel de 55 gènes chez 48 sujets (30 femmes HetAT, 12 femmes non-HetAT et 3 patients A-T) ayant développé un cancer a permis l'identification de cinq variants rares de type « perte de fonction » chez cinq patients provenant de familles A-T différentes, 13 variants faux-sens rares classés comme probablement délétères ou VSI par les différents outils de prédiction cités précédemment (Tableau 10).

Tableau 10 : Liste des variants identifiés chez les individus séquencés avec le panel de 55 gènes.

Individu	Mutation ATM familiale	Mutation ATM confirmée	Cancer	Autre gène altéré	Variants	Effet sur la protéine	Référence dbSNP	Fréquence EVS (%)	Grade Align-GVGD	Classement selon ClinVar	SIFT	PolyPhen
AT023006	c.5644C>T	oui	Sein	BRCA2	c.2944A>C	p.Ile982Leu	rs28897717	-	C0	VSI pour les cancers du sein et de l'ovaire	Toléré	Bénin
AT082001	c.3801delG	oui	Sein	FANCF	c.1087C>T	p.Gln363*	rs201285915	0.01	-	VSI ¹	-	-
AT107020	c.5763-2A>T	oui	Sein	MUTYH	c.270C>A	p.Tyr90*	rs121908380	0.2	-	-	-	-
AT107021	c.5763-2A>T	oui	Sein	MUTYH	c.270C>A	p.Tyr90*	rs121908380	0.2	-	-	-	-
AT107021	c.5763-2A>T	oui	Sein	BRIP1	c.550G>T	p.Asp184Thr	rs201047375	0.01	-	VSI pour les cancers héréditaires	Délétère	Probablement délétère
AT149008	c.8124TT>A	oui	Sein	MSH6	c.3727A>T	p.Thr1243Ser	rs147453999	0.02	C55	VSI pour les cancers héréditaires	Délétère	Possiblement délétère
AT047002	c.73-2A>G	oui	Pancréas	CDH1	c.1774AG>A	p.Ala592Thr	rs35187787	0.5	-	Pas de consensus ²	Délétère	Possiblement délétère
AT186002	c.2922-2A>G c.3248A>G	oui	Estomac	FANCM	c.171G>C	p.Leu57Phe	rs142007602	0.2	-	VSI pour l'anémie de Fanconi	Délétère	Bénin
AT211003	c.6680G>T c.8283_8284delTC	oui	Maladies immunoprolifératives malignes	PALB2	c.2435C>T	p.Pro812Leu	rs774502617	-	C0	VSI ¹	Toléré	Bénin
AT211003	c.6680G>T c.8283_8284delTC	oui	Maladies immunoprolifératives malignes	MET	c.3253C>A	p.Gln1085Lys	rs45628136	-	-	VSI pour le carcinome rénal	Délétère	Possiblement délétère
AT108041	non-HetAT	-	Sein	FANCM	c.1576C>G	p.Leu526Val	rs144215747	0.1	-	VSI pour l'anémie de Fanconi	Délétère	Bénin
AT138022	non-HetAT	-	Sein	BRCA1¶	c.1016dupA	p.Val340Glyfs*6	rs80357569	-	-	Pathogène pour les cancers du sein et de l'ovaire	-	-
AT096024	non-HetAT	-	Utérus/Sein	ATM	c.7658C>A	p.Pro2553His	rs864622368	-	C65	VSI pour l'A-T	Délétère	Probablement délétère
AT141010	non-HetAT	-	Lymphome non-Hodgkinien	FANCL	c.1114insATTA	p.Thr372Asnfs*13	rs759217526	0.2	-	Pas de consensus ³	-	-
AT242004	non-HetAT	-	Ovaires	MSH6	c.3820G>T	p.Glu1274*	-	-	-	-	-	-
AT273025	non-HetAT	-	Rein	RINT1	c.1985T>C	p.Ley662Ser	rs139459919	0.02	-	VSI ¹	Délétère	Probablement délétère
AT404002	non-HetAT	-	Poumon	ATM	c.4324T>C	p.Tyr1442His	rs201666889	0.02	C65	Bénin pour la prédisposition au cancer VSI pour l'A-T	Toléré	Probablement délétère
AT404002	non-HetAT	-	Poumon	FANCM	c.527C>T	p.Thr176Ile	rs77374493	0.4	-	Pas de consensus ⁴	Toléré	Bénin
AT404002	non-HetAT	-	Poumon	BRIP1	c.2344A>G	p.Ile782Val	rs142806416	-	-	VSI pour le cancer du sein familial	Toléré	Bénin

VSI : variant de signification clinique inconnue. (-) Pas d'information sur l'effet de la protéine. ¶ Mutation déjà identifiée dans cette famille. ¹Variant considéré comme VSI dans le cadre de tests cliniques. ²Variant rapporté dans ClinVar (3 fois comme « bénin », 4 fois probablement bénin et une fois VSI dans le cadre de prédisposition familiale au cancer). ³Variant classé dans le cadre de l'anémie de Fanconi 2 fois comme pathogène, une fois comme bénin et une fois comme VSI. ⁴Variant classé une fois comme VSI et une fois comme bénin dans le cadre de l'anémie de Fanconi.

3.4.1 Variants « perte de fonction » identifiés

Le reséquençage du panel de 55 gènes a conduit à l'identification de trois variants « perte de fonction » situés dans les gènes *FANCF* (patiente HetAT AT082001 ayant développé un cancer du sein), *MUTYH* (patientes HetAT AT107020 et AT107021 ayant développé toutes deux cancer du sein) et *MSH6* (patiente non-HetAT AT242004 ayant développé un cancer de l'ovaire).

La patiente HetAT AT082001 portant le variant d'*ATM* pathogène pour l'A-T (c.3802delG ; p.Val1268X) est également porteuse hétérozygote du variant non-sens c.C1087T (p.Gln363X) dans le gène *FANCF* (Tableau 10). Cette variation n'est ni rapportée dans la base de données LOVD (<http://databases.lovd.nl>), ni dans les bases ClinVar (www.ncbi.nlm.nih.gov/clinvar) ou ClinGen (www.clinicalgenome.org). *FANCF*, comme les autres protéines FANC est impliquée dans la réparation de l'ADN via l'ubiquitination de *BRCA2*, mais les variants localisés dans le gène *FANCF* ne semblent cependant pas associé au risque de cancer du sein¹⁹¹.

Dans la famille AT107 (Figure 18), les deux sœurs HetAT AT107020 et AT107021 portent à l'état hétérozygote, en plus de la mutation familiale d'*ATM* (c.5763-2A>T), un variant non-sens dans le gène *MUTYH*, (c.C270A ; p.Tyr90X) (Tableau 10) non répertorié dans les bases ClinVar et LOVD. La protéine *MUTYH* intervient dans la réparation des dommages à l'ADN causés par le stress oxydatif. Des mutations bi-alléliques dans le gène *MUTYH* sont associées aux formes de polypose adénomateuse familiale (PAF) (ou en anglais « MAP » pour « *MUTYH-associated polyposis* »). Dans ces familles, plusieurs cas de cancer avec polyposes ainsi que des cas multiples de cancer colorectal sont généralement diagnostiqués¹⁹². Une association entre cancer du sein et mutation mono-allélique du gène *MUTYH* a été montrée dans des familles prédisposées aux PAF^{193,194} mais une étude cas/témoins menées chez des cas de cancer du sein sans mutation dans les gènes *BRCA1* et *BRCA2* a montré que le risque de développer ce cancer n'était pas augmenté chez les porteurs hétérozygotes d'une mutation dans *MUTYH*¹⁹⁵. Il semble donc peu probable que ce variant soit responsable de la prédisposition au cancer du sein dans la famille AT107.

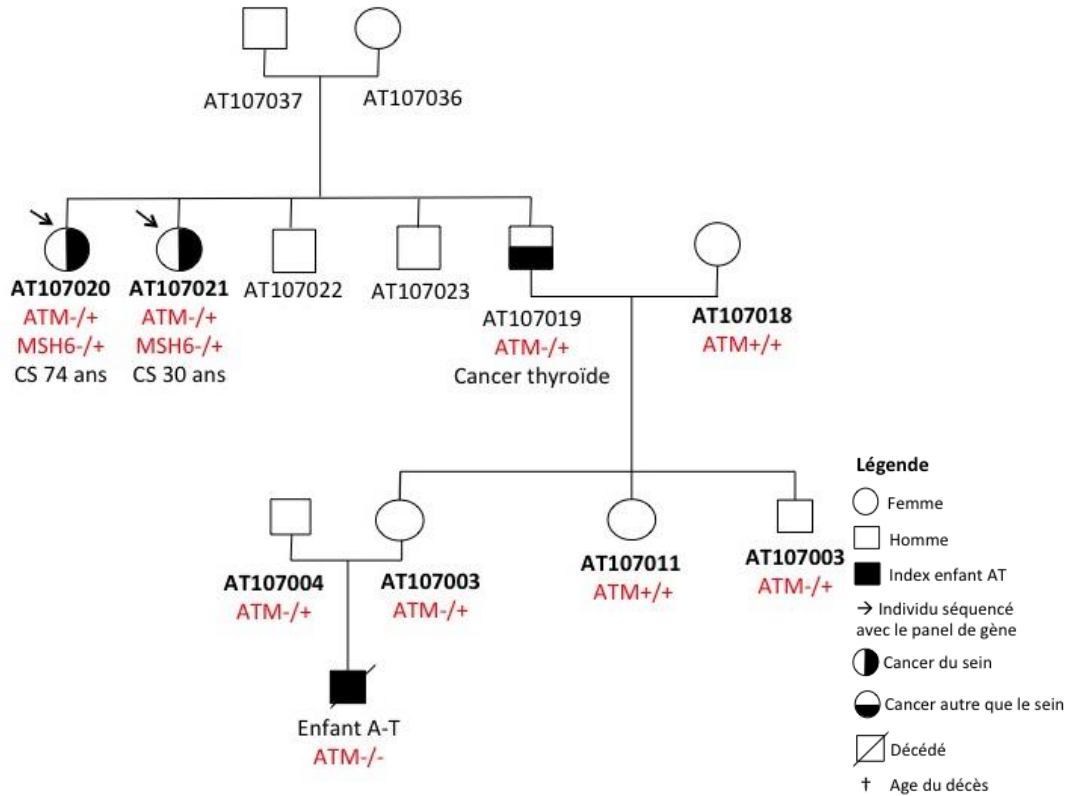


Figure 18 : Arbre généalogique simplifié de la famille AT107.

Les individus en gras sont les participants à CoF-AT ou Retro-AT pour lesquels un échantillon d'ADN est disponible. ATM-/+: porteur hétérozygote de la mutation familiale ATM, ATM+/+: non-porteur de la mutation familiale ATM. Les flèches indiquent les individus séquencés pour les 55 gènes du panel.

Enfin, la patiente non-HetAT AT242004 ayant développé un cancer de l'ovaire porte un variant non-sens (c.3820G>T ; p.Glu1274*) sur le gène *MSH6* (Tableau 10). Le gène *MSH6* code pour une protéine impliquée dans la détection des bases non-appariées de l'ADN et est impliqué dans les formes familiales HNPCC prédisposant aux cancers colorectaux et de l'endomètre^{196, 197}. Les tumeurs liées à une mutation de ce gène étant caractérisées par une instabilité des microsatellites¹⁹⁸, j'ai donc recherché cette instabilité dans l'ADN tumoral de la patiente au moyen de 7 marqueurs microsatellites localisés à différentes positions du génome : BAT-25, BAT-26, NR-21, NR-24 et MONO-27 (the MSI Analysis System, v.1.2 Promega). Les résultats montrent une instabilité des marqueurs BAT-25 et MONO-27. Une instabilité d'au moins 2 des 5 marqueurs étant considérée comme une instabilité forte¹⁹⁹, ce résultat pourrait être une indication quant à la pathogénicité du variant c.3820G>T dans la famille AT242 (Figure 19).

Six autres membres de la famille AT242 ont développé un cancer (Figure 20), dont trois femmes participant à CoF-AT (AT242001, AT242003 et AT242005). La recherche ciblée du variant *MSH6* par séquençage de Sanger montre qu'aucune de ces participantes n'est porteuse du variant c.3028G>T (Figures 21).

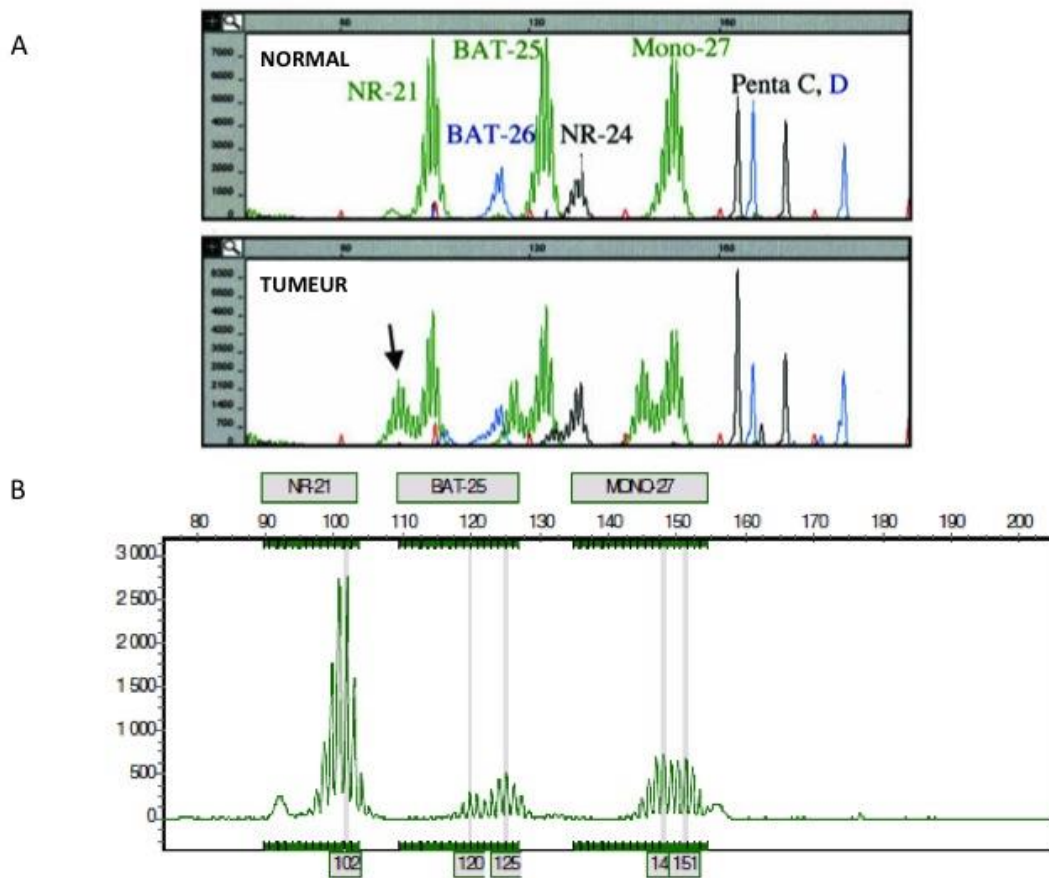


Figure 19 : Résultats de l'analyse de l'instabilité des microsatellites dans l'ADN tumoral de la patiente AT242004.
A. Profil attendu d'un échantillon d'ADN normal sans instabilité des microsatellites et d'un échantillon d'ADN tumoral présentant une instabilité des cinq marqueurs microsatellites. **B.** Les marqueurs microsatellites BAT-25 et MONO-27 présentent un profil similaire à celui de la tumeur du cadre A, caractéristique d'une instabilité liée à une mutation dans le gène *MSH6*.

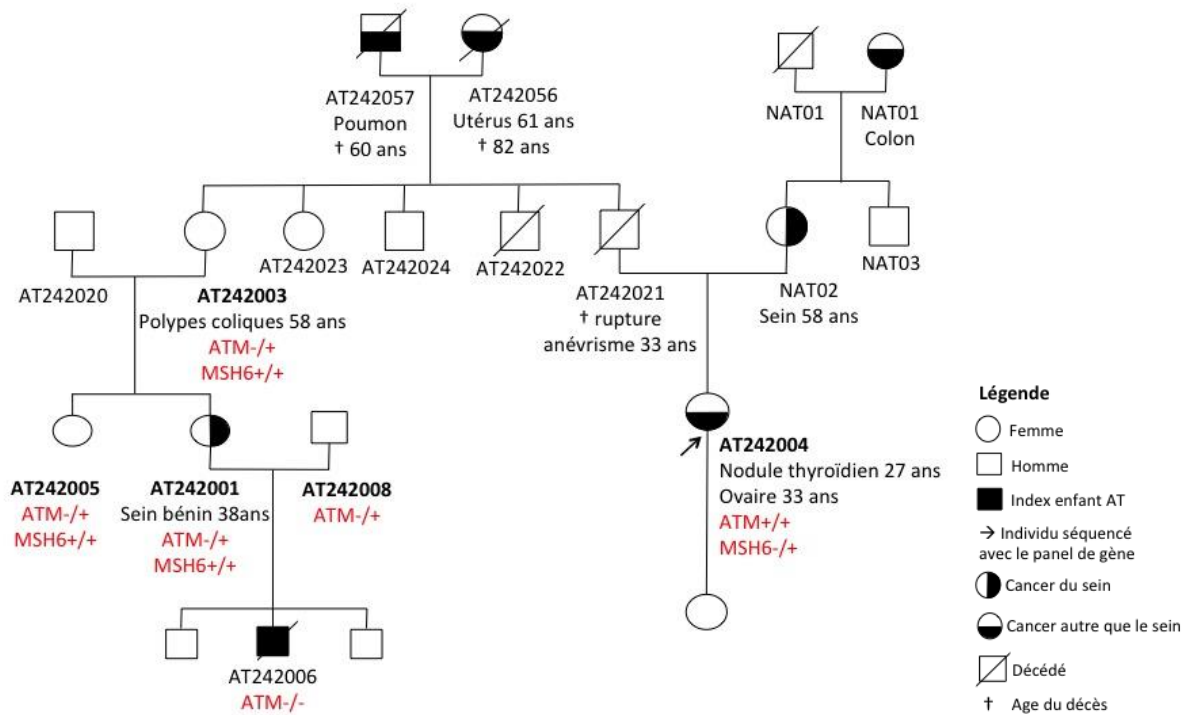


Figure 20 : Arbre généalogique simplifié de la famille AT242.

L'enfant A-T (AT242006) est l'enfant ayant permis l'identification de la famille. Les individus en gras sont les individus pour lesquels un échantillon d'ADN est disponible. Les résultats des analyses moléculaires pour les gènes *ATM* et *MSH6* sont indiqués en rouge : *ATM*+/+ : non-porteur de la mutation familiale pour *ATM* ; *ATM*-/+ : porteur hétérozygote de la mutation familiale pour *ATM* ; *ATM*-/- : porteur homozygote composite des mutations maternelle et paternelle d'*ATM* ; *MSH6*+/+ : non-porteur de la mutation c.3820G>T dans *MSH6* ; *ATM*-/+ : porteur hétérozygote de la mutation c.3820G>T dans *MSH6*.

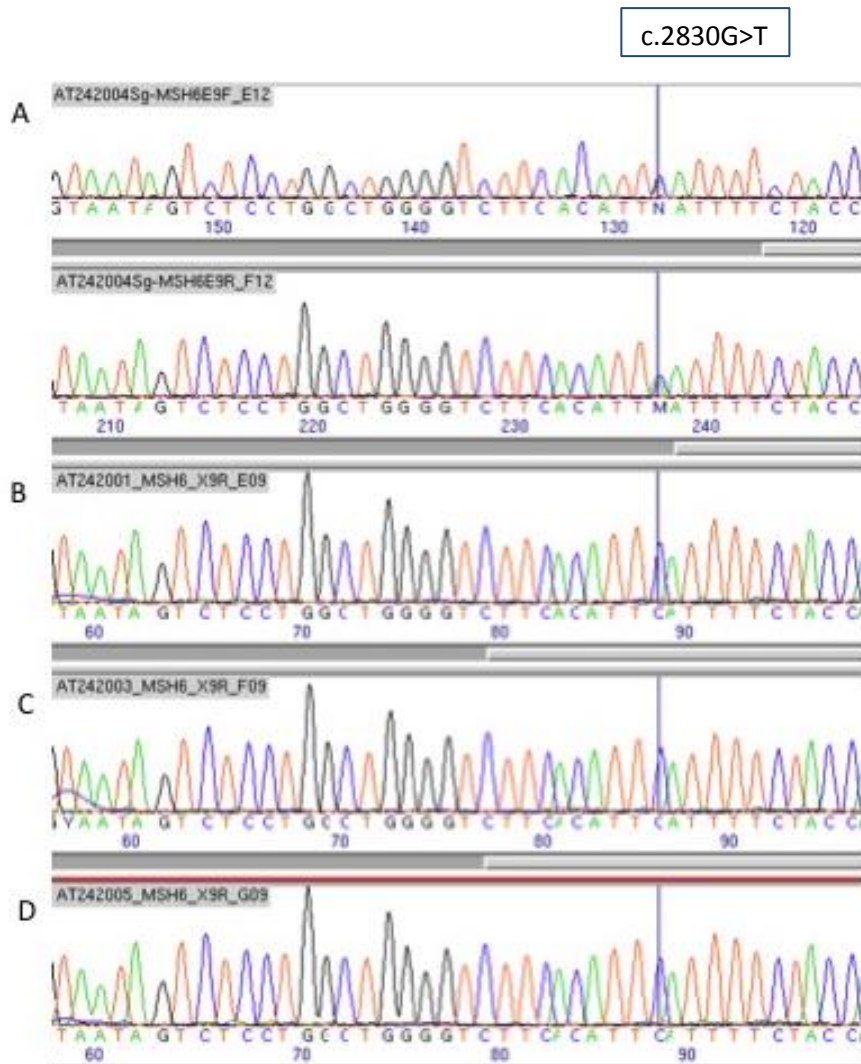


Figure 21 : Résultats du séquençage ciblant le variant c.3028G>T de *MSH6* pour les participantes de la famille AT242.

A. Chromatogrammes confirmant le variant identifié par séquençage NGS du panel de gènes chez la patiente AT242004. La position de la mutation est indiquée par le trait vertical bleu. **B, C, D.** Chromatogrammes correspondant à la recherche du variant chez les autres membres de la famille (AT242001, AT242005 et AT242003).

L'ensemble des observations concernant la famille AT242 ouvre sur différentes hypothèses :

- le variant du gène *MSH6* est transmis par la branche maternelle de la participante AT242004 chez qui le variant a été découvert,
- la participante AT242004 a hérité du variant *MSH6* de son père (AT242021),
- il s'agit d'un variant *de novo*.

Malheureusement nous ne pouvons confirmer aucune de ces hypothèses car nous n'avons pas accès au prélèvement sanguin d'autres membres de cette famille. La branche maternelle de la participante AT242004 ne peut être incluse dans l'étude CoF-AT car elle ne possède pas de lien avec l'enfant atteint d'A-T et les autres membres de la branche paternelle ont soit refusé de participer à CoF-AT ou ne sont pas éligibles pour cette étude.

Le fait qu'un VSI de *MSH6* ségrége dans la famille AT242 devra donc être pris en compte dans les analyses futures visant à estimer les risques de cancer pour les porteurs de la mutation familiale

d'*ATM* dans cette famille.

3.4.2 Variants faux-sens de signification clinique inconnue.

A l'issue du séquençage du panel de gènes chez les sujets atteints de cancer dans les familles A-T étudiées, trois variants faux-sens classés « délétères » par SIFT, « probablement délétères » par PolyPhen et avec les gardes les plus sévères C65 et C55 par Align-GVGD ont retenus notre attention. L'impact des variants sur la protéine définis à partir de ces trois outils *in silico* concordent, sauf pour le variant c.4324T>C (p.Tyr1442His) d'*ATM* classé comme « toléré » par SIFT.

La patiente AT149008 porte, en plus de la mutation familiale *ATM*, un variant faux-sens dans le gène *MSH6* (c.A3727T, pThr1243Ser) classé C55 par Align-GVGD et rapporté comme VSI pour les cancers héréditaires dans ClinVar. Les données cliniques et familiales de cette famille ne mettent pas en avant une prédisposition génétique au cancer colorectal, suggérant que le variant n'intervient pas dans le développement de la maladie ou que sa pénétrance est incomplète.

Deux variants faux-sens d'*ATM* ont été identifiés chez les patientes AT096024 (c.C7658A ; p.Pro2553His) et AT404002 (c.4324T>C ; p.Tyr1442His) non porteuses de la mutation familiale responsable de l'A-T dans cette famille (Tableau 10). Seule l'analyse de la ségrégation du variant c.7658C>A a pu être réalisée dans la famille AT096. Le séquençage ciblé du variant chez six membres de la famille révèle que trois d'entre eux sont porteurs hétérozygotes (AT096018, AT096005 et AT096002) (Figure 22), dont un patient atteint d'un cancer de la prostate. Bien que les résultats de séquençage combinés aux données familiales et cliniques ne permettent pas de conclure sur l'implication du variant dans le risque de cancer du sein, ce variant rapporté comme VSI pour l'A-T dans ClinVar ne semble pas être impliqué dans ce syndrome d'après les données dont nous disposons à ce jour. En effet, le variant est retrouvé chez le père de l'enfant atteint d'A-T (AT096005), à la fois porteur hétérozygote de la mutation délétère d'*ATM* responsable de la maladie (p.Val1268X) et du variant faux-sens c.7658C>A.

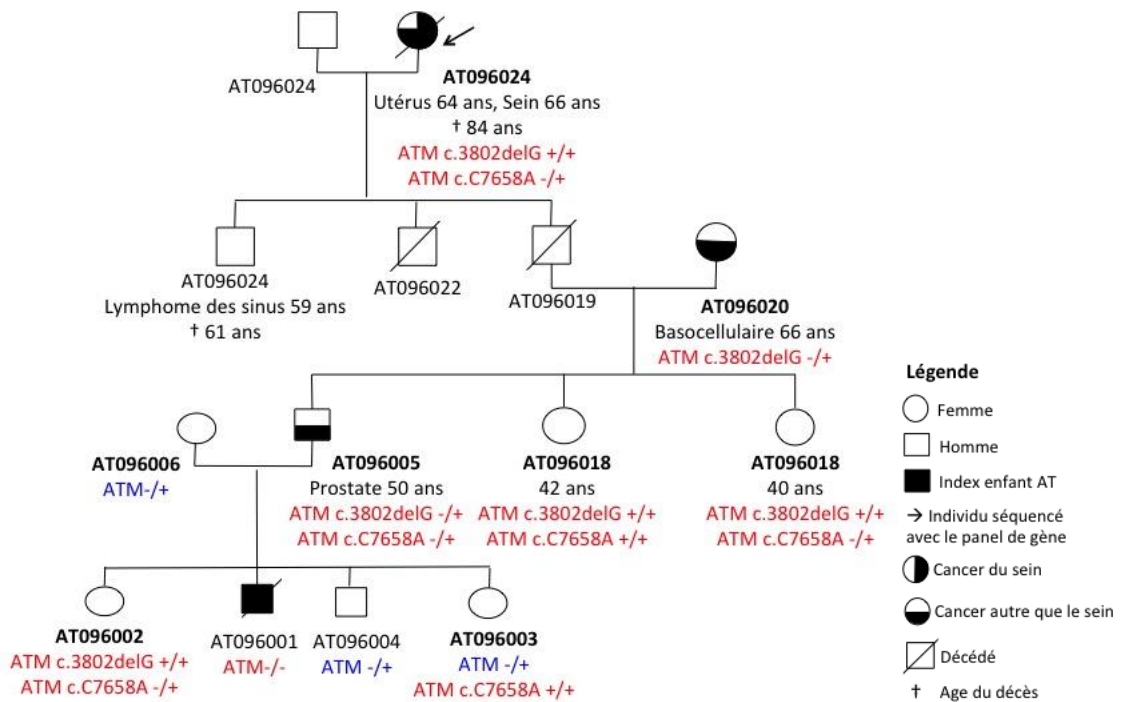


Figure 22 : Arbre généalogique simplifié de la famille AT096.

L'enfant A-T (AT096001) est l'enfant ayant permis l'identification de la famille. Les identifiants des individus en gras indiquent les participants pour lesquels un échantillon d'ADN est disponible. Les variants d'ATM qui ségrègent dans la famille sont indiqués pour chaque individu lorsque que la variation a été recherchée. ATM +/+ : non-porteur de la mutation pour ATM ; ATM -/+ : porteur hétérozygote de la mutation pour ATM ; ATM -/- : porteur homozygote composite des mutations maternelles et paternelles dans le ATM.

En conclusion, une faible proportion des familles A-T étudiées ségrègent en plus d'ATM un variant rare délétère dans un autre gène de prédisposition au cancer. Bien que trois participantes de CoF-AT portent un variant tronquant associé à une perte de fonction de *MSH6*, *FANCM* ou *MUTYH*, les éléments dont nous disposons sont insuffisants pour établir clairement un lien entre le variant et le développement du cancer. Dans la suite de mes travaux portant sur la recherche de facteurs modificateurs du risque de cancer dans les familles A-T, j'ai décidé d'exclure les femmes porteuses d'une mutation dans les gènes *BRCA1* ou *BRCA2* et la patiente porteuse du variant délétère de *MSH6*.

4. Les télomères

Après m'être concentrée sur la recherche de variants rares dans d'autres gènes de prédisposition au cancer, je me suis intéressée à la longueur des télomères. Mon intérêt s'est porté sur cet autre facteur constitutionnel car ATM possède un rôle direct dans la maintenance de l'élongation des télomères. Avant d'exposer les résultats issus de ces travaux, nous allons faire un bref rappel sur les télomères.

L'extrémité des chromosomes est constituée de séquences nucléotidiques répétées $(TTAGGG)_n$ ²⁰⁰ protégées par un complexe protéique formé par les protéines shelterin²⁰¹. Cet ensemble forme les télomères, un complexe nucléoprotéique dynamique qui mesure à la naissance entre 10 et 15 kb²⁰² et qui se raccourcit à chaque division cellulaire. Les télomères se raccourcissent progressivement jusqu'à une longueur critique qui entraîne la cellule en phase de sénescence ou à l'apoptose²⁰³ (Figure 23). En plus de leur fonction d'horloge biologique de la cellule, les télomères veillent au maintien de l'intégrité du génome en empêchant la dégradation de l'ADN par les nucléases ainsi que les fusions ou les réarrangements des extrémités des chromosomes^{202, 204, 205}.

La longueur moyenne des télomères d'un individu est donc considérée comme un bio-marqueur du vieillissement cellulaire.

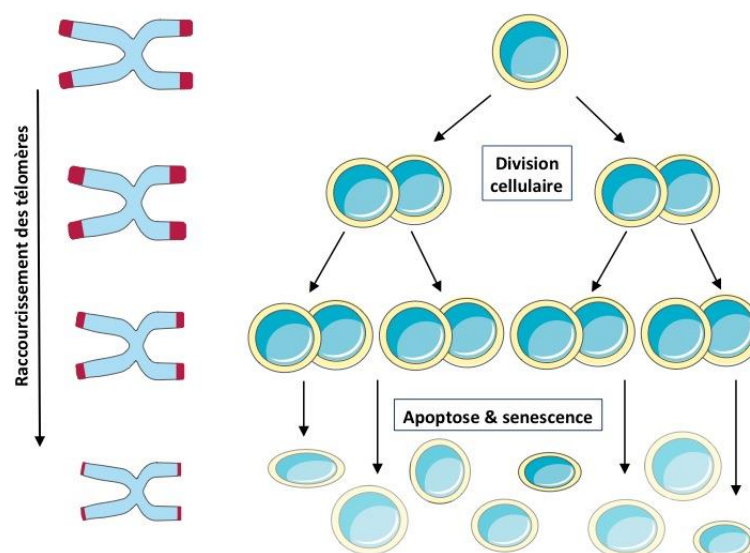


Figure 23 : Raccourcissement des télomères au cours de la division cellulaire (d'après S. Leighton, The Scientist, 2012).

4.1 Structure de l'ADN télomérique

Les répétitions des nucléotides TTAGGG forment une séquence d'ADN double-brin se terminant en

ADN 3'-simple brin riche en G sur une longueur de 50 à 500 nucléotides ²⁰². En amont de la partie double-brin des télomères se trouve une séquence d'ADN spécifique subtélomérique et une séquence d'ADN dégénérée de séquences nucléotidiques TTAGGG (Figure 24.A).

Griffith et *al.* ont montré que l'ADN 3'-simple brin a la capacité d'envahir l'ADN télomérique double-brin pour donner une conformation en boucle dite en « t-loop » ²⁰⁶ (Figure 24.B). Cette structure permet à la cellule de distinguer l'ADN simple brin situé à l'extrémité des chromosomes des cassures d'ADN survenant sur le reste du génome et nécessitant l'intervention des mécanismes de réparation

207 .

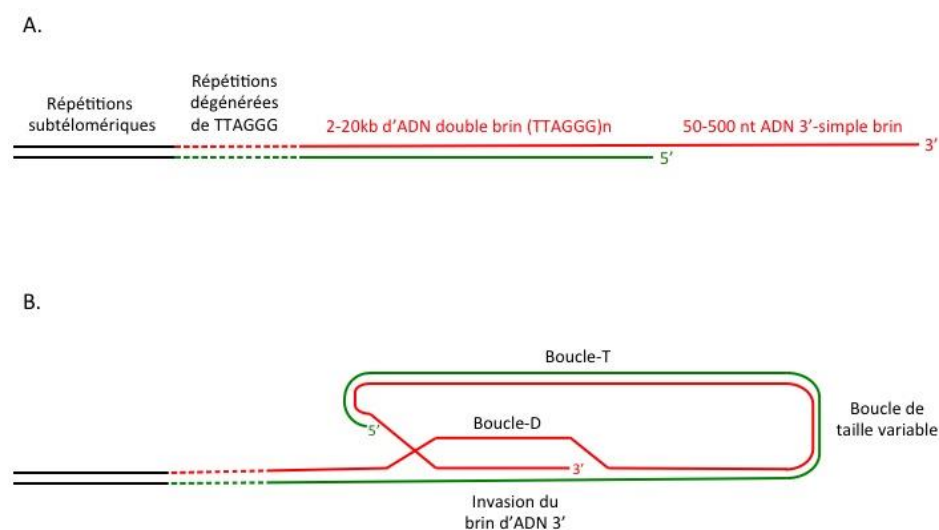


Figure 24 : Représentation de la structure des télomères dite en "t-loop".

a. Structure des extrémités des chromosomes. b. Schéma de la formation de la boucle dite en t-loop (d'après W. Palm, Annual Reviews Genetics, 2008 ²⁰²).

4.2 Rôle des protéines shelterin

Les protéines shelterin forment un complexe de six protéines connectées les unes aux autres (Figure 25) :

- TRF1 (*TTAGGG Repeat Factor 1*) fixe l'ADN double-brin spécifiquement au niveau des répétitions de TTAGGG et a pour rôle de réguler l'élongation des télomères en empêchant la fixation de la télomérase ^{208, 209} ;
- TRF2 (*TTAGGG Repeat Factor 2*) fixe l'ADN double-brin spécifiquement au niveau des répétitions de TTAGGG ²⁰² et a la capacité d'entraîner la formation de structure en t-loop ²¹⁰ ;
- RAP1 (*Repressor/Activator Protein 1*) : forme un complexe avec TRF2 ²¹¹ pour stabiliser la fixation de cette dernière sur les télomères ^{212, 213} ;

- POT1 (*Protection Of Telomeres protein 1*) : fixe les répétitions de TTAGGG au niveau de l'extrémité 3' simple-brin de l'ADN^{214, 215} ou au niveau de l'envahissement de l'ADN double-brin pour former la structure en t-loop ;
- TIN2 (*TRF1 - and - TRF2 Interacting Nuclear protein 2*) forme un pont entre les protéines TRF1 et TRF2^{216, 217} et sert également de connecteur entre TRF1 et POT1. TIN2 permet donc de faire le lien entre les protéines fixant l'ADN double-brin et POT1 liée à l'ADN simple-brin²¹⁸ ;
- TPP1 (*TriPeptidyl Peptidase 1*) permet de recruter POT1 au niveau des télomères²¹⁸⁻²²⁰ et de lier POT1 et TIN2²²¹.

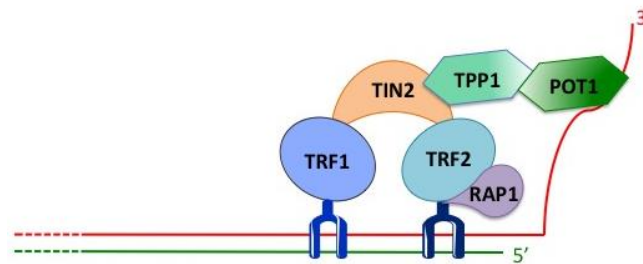


Figure 25 : Le complexe des protéines shelterins.
(d'après T. de Lange, Science, 2009²²²).

En plus de leur rôle dans la conformation des télomères les protéines TRF1, TRF2, RAP1 et POT1 inhibent également la voie de réparation des dommages à l'ADN par recombinaison-homologue en inactivant les protéines ATM et ATR^{223, 224}.

Les protéines shelterins permettent donc de protéger et de maintenir la stabilité du génome en intervenant dans la formation de la structure en t-loop, en servant de régulateur à l'élongation des télomères et en inhibant les voies de réparation liées aux cassures de l'ADN^{202, 225, 226}.

4.3 Régulation de la longueur des télomères

4.3.1 Rôle de la télomérase

La perte d'un fragment télomérique à chaque division cellulaire est compensée par l'élongation des télomères par la télomérase, un complexe enzymatique composé de la sous-unité protéique TERT (*Telomerase Reverse Transcriptase*), de la sous-unité nucléotidique hTR (ou TERC) (*Telomerase RNA Component*) et de cinq protéines contribuant à la stabilité du complexe et à ses déplacements dans le noyau²²⁷⁻²²⁹ : DKC (*dyskerin*), NPH2 (*non-histone protein 2*), NOP10 (*nucleolar protein 10*), TCAB1 (*Telomerase Cajal body protein 1*) et GAR (Figure 26).

Chez l'adulte la prolifération des cellules n'étant pas une nécessité pour la majorité des tissus, la télomérase reste inactivée. On observe cependant une exception au niveau des cellules souches

notamment dans les cellules hématopoïétiques.

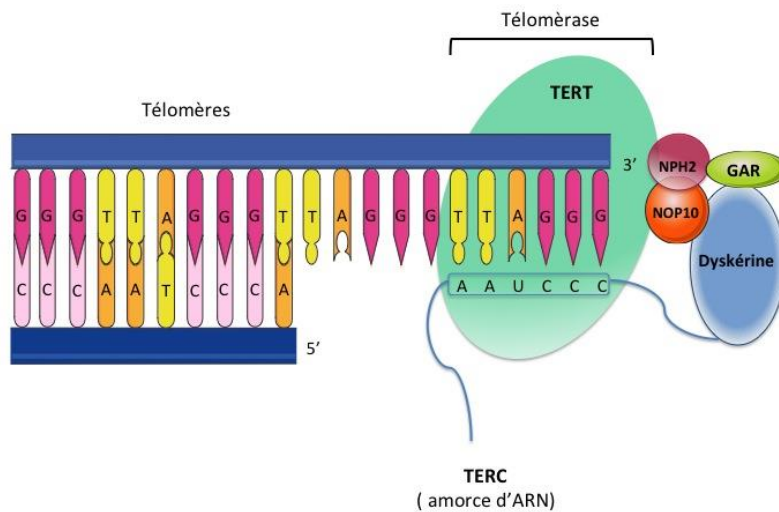


Figure 26 : Protéines et structure de la télomérase.
(d'après Calado et Young, 2009²³⁰).

4.3.2 Elongation des télomères

La sous-unité nucléotidique TERC possédant une séquence d'ARN complémentaire à l'ADN se fixe la première sur l'ADN via une interaction avec la protéine shelterin TTP1. Dans un deuxième temps la sous-unité TERT se lie à TERC afin d'ajouter progressivement les nucléotides nécessaires à l'élongation des télomères. Le recrutement de la télomérase au niveau des télomères est régulé par les protéines shelterin TRF1, TFR2, TTP1 et RAP1. En effet, la télomérase est recrutée lorsque la cellule présente un faible nombre d'unités répétées de TTAGG et donc que la quantité de protéine TRF1, TRF2, TTP1 et RAP1 au niveau des télomères est faible^{202, 209, 217}.

Récemment, il a été démontré que le recrutement de la télomérase est également régulée via les protéines de la réparation ATM et ATR. Ce processus est détaillé dans le premier article émanant de mes travaux de thèse portant sur la longueur des télomères chez les individus porteur d'une copie mutée d'*ATM* (pages 106).

4.4 Télomères, vieillissement et cancer

4.4.1 Syndromes associés au vieillissement

De part leur fonction biologique les télomères sont associés au phénomène naturel de vieillissement mais des défauts chez les différents acteurs de la maintenance des télomères sont responsables des syndromes de vieillissement prématuré, comme par exemple la dyskératose congénitale²³¹. Ce syndrome caractérisé par des télomères plus courts chez les jeunes malades^{232, 233} est causé par des mutations dans les gènes *TERC*²³⁴⁻²³⁶, *TERT*²³⁷⁻²³⁹ et *TIN2*^{240, 241}. On trouve également des mutations

dans les gènes *TERT* et *TERC* chez les individus atteints de fibrose pulmonaire²⁴²⁻²⁴⁴. Les deux formes de progéria les plus connues, le syndrome de Werner et le syndrome de Hutchinson-Gilford, sont quant à eux associés respectivement à des mutations dans les gènes *WRN*^{245, 246} et *LMNA*²⁴⁷ impliqués dans la réparation de l'ADN.

4.4.2 Télomères et cancer

Dans les cellules normales, la longueur des télomères sert de point de contrôle au passage en phase de senescence par une activation en cascade des protéines ATM et ATR²⁴⁸, puis des protéines p53 et RB. Cependant, certaines cellules accumulant les modifications oncogéniques contournent ce point de contrôle et entrent dans une phase de crise. Durant cette phase les cellules « luttent » contre une forte instabilité génomique, due à une perte des télomères sur certains chromosomes²²⁸, et qui se caractérise par des pertes d'hétérozygotie, des translocations, des amplifications, une tétraploïdisation des cellules, une chromothripsie¹, une kataegis² ou encore des fusions de chromosomes²⁴⁹⁻²⁵⁴. Lors de cette phase de crise, la réactivation de la télomérase entraîne l'immortalisation des cellules et la formation d'un clone de cellules tumorales prêt à proliférer.

Il a été montré que dans 80% des cancers les cellules sont immortalisées par l'activation continue de la télomérase²⁵⁵. Des études sur les formes familiales ou sporadiques de mélanome, de gliome, sur le médulloblastome ou encore le cancer du foie ont montré que de l'activation de la télomérase était induite par des mutations dans le promoteur du gène *TERT*²⁵⁶⁻²⁶⁰. Par exemple, la présence dans le promoteur du gène *TERT* des variants c.-57A>C, c.-124C>T et c.-146C>T en amont du site initial de transcription crée de nouveaux sites de fixation aux facteurs de transcription ETS, ce qui augmentent la production d'ARNm ainsi que l'expression des protéines TERT²⁶¹. Ces différentes observations montrent que le maintien de l'activité de la télomérase perturbe le rôle des télomères en tant que suppresseur de tumeur.

Des mutations constitutionnelles dans le gène *POT1* codant la protéine du même nom ont été mises en évidence dans des cas familiaux de mélanome^{262, 263}. Les porteurs de mutations dans ce gène présentent des télomères plus longs que les individus non porteurs d'une mutation dans ce gène. Par

¹ *Chromothripsie* : Réarrangement aléatoire de fragments chromosomiques au cours d'un événement cellulaire. Lors de la division de cellules en phase de crise deux chromosomes ayant fusionné forment un pont entre les cellules filles, ce qui empêche leur division. La rupture du pont est rendue possible par l'action de la protéine TREX1. Lors de ce processus des fragments chromosomiques sont produits aléatoirement et se dispersent dans l'une ou l'autre des deux cellules filles. Les fragments sont soit perdus soit réparés de façon aléatoire sur les chromosomes.

² *Kataegis* : Modèle d'hyper-mutation C>T et C>G localisée au niveau de régions génomiques réarrangées.

ailleurs, des télomères longs ont également été associés à un risque accru de leucémie lymphocytaire chronique ^{264, 265}, de cancer du poumon ²⁶⁶ ou de mélanome ²⁶². Deux études récentes ont montré que les télomères longs chez les porteurs de mutations de *POT1* seraient probablement dus à l'incapacité de la protéine POT1 à se fixer sur l'ADN simple-brin et à inhiber la télomérase. La mutation dans le gène *POT1* repousserait à la fois la phase de senescence via l'activation de la télomérase tout en augmentant l'instabilité génomique, et donc le risque de mélanome.

4.5 Problématique posée dans le cadre de l'étude sur les familles A-T

Différentes études confirment le rôle essentiel des télomères comme suppresseur de tumeur et montrent que le lien entre longueur des télomères et cancer dépend d'un équilibre entre télomères longs et télomères courts. D'autre part, les données sur *ATM* suggérant que la protéine est impliquée dans la régulation de l'élongation des télomères, nous avons émis l'hypothèse que la longueur des télomères varie en fonction du statut mutationnel d'*ATM* et que le risque de cancer dans les familles liées à *ATM* est modulé par longueur des télomères.

4.6 Mesure de la longueur des télomères chez l'ensemble des participants

4.6.1 Choix de la technique

Différentes techniques peuvent être utilisées pour estimer la longueur des télomères (LT) : le transfert d'ADN communément nommé « Southern blot » ^{267, 268}, la PCR quantitative (q-PCR) ^{269, 270} et l'hybridation quantitative *in situ* en fluorescence (Q-FISH pour *Quantitative Fluorescence in situ Hybridization*) ^{271, 272}.

La technique de Southern blot consiste à séparer des fragments d'ADN de différentes tailles en les faisant migrer sur un gel d'agarose. Cette technique nécessite d'extraire l'ADN à partir de cellules ou de tissus, de digérer d'ADN par des enzymes de restriction qui digèrent à la fois les parties télomériques et subtélomériques, d'évaluer la qualité de l'ADN digéré par une q-PCR, réaliser une électrophorèse sur gel d'agarose pour ensuite transférer l'ADN sur une membrane sur laquelle les sondes d'ARN fluorescentes seront hybridées et révélées ²⁶⁸. La mesure des télomères par cette technique est considérée comme la technique de référence car elle permet de mesurer la taille réelle des télomères.

La technique de Q-FISH repose sur l'hybridation au niveau des télomères de sondes nucléotidiques fluorescentes dans les cellules ou sur une coupe de tissu.

La q-PCR, qu'elle soit simplexe (amplification d'un seul fragment d'ADN) ou multiplexe (amplification simultanée de multiples fragments), consiste à amplifier par PCR les télomères et une copie unique d'un gène de référence afin de calculer un ratio télomères/gène de référence correspondant à une longueur relative de télomères.

Les avantages et les inconvénients de chacune de ces techniques sont présentés dans le tableau 11.

Tableau 11 : Comparaison des différentes techniques utilisées dans la littérature pour mesurer la longueur des télomères.

Technique	Références	Matériel analysé	Principe	Avantages	Inconvénients
Southern blot	- Kimura M et <i>al.</i> , Nature Protoc, 2010 ²⁶⁸ - Kimura M et <i>al.</i> , Nucleic Acids Res, 2011 ²⁶⁷	ADN	-Isolation des régions télomériques par digestion par des enzymes de restriction - Migration sur gel d'agarose -Hybridation de l'ADN télomérique avec des sondes d'ARN fluorescentes.	Mesure la longueur réelle (kb) des télomères pour chaque échantillon.	- Protocole s'étalant sur plusieurs jours et coûteux, ne peut être automatisé. - Grande quantité d'ADN requise (au moins 3ug). - Clivage aléatoire de l'ADN par les enzymes dans les régions télomériques et subtélomériques entraînant une variation de la LT entre 2 réactions enzymatique
q-FISH	- Ferlicot S <i>al.</i> , J Patho, 2003 ²⁷¹ - Ourliac-Garnier I <i>al.</i> , Methods Mol Bio, 2003 ²⁷²	Tissu, Cellules	Hybridation de sondes nucléotidiques fluorescentes sur des coupes de tissu ou des cellules fixées.	- Intensité du signal proportionnelle au nombre de répétitions des télomères. - Mesure réelle de la LT pour chaque chromosome.	- Protocole nécessitant plusieurs étapes - Pas d'automatisation possible. - Non applicable à de grandes séries d'échantillons.
q-PCR simplex	- Cawthon RM <i>al.</i> , Nucleic Acids Res, 2002 ²⁶⁹	ADN	Amplification par PCR en temps réel des télomères et d'un gène de référence.	- Protocole rapide (quelques heures). - Possibilité d'amplifier une centaine d'échantillons en même temps (plaques 96-puits). - Peu de matériel de départ (10ng) - Faible coût (0.25€/échantillon)	- Estimation d'une longueur relative des télomères. - Risque de biais dans les résultats causés par des quantités différentes d'ADN entre la réaction d'amplification du gène de référence et celle des télomères.
MM QPCR	- Cawthon RM <i>al.</i> , Nucleic Acids Res, 2009 ²⁷⁰	ADN	Amplification simultanée par PCR en temps réel des télomères et d'un gène de référence dans une même réaction.	- Protocole rapide (quelques heures). - Possibilité d'amplifier une centaine d'échantillons en même temps. - Pas de biais de puits, de manipulation ou de différence de quantité d'ADN pour chaque échantillon.	- Estimation de la longueur relative des télomères.

4.6.2 Mise en place du protocole de q-PCR

Pour notre étude nécessitant l'analyse de plusieurs centaines d'échantillons, j'ai choisi d'optimiser au laboratoire la méthode de q-PCR pour estimer la longueur relative des télomères chez les participantes. La méthode de q-PCR nécessite d'établir une gamme d'étalonnage pour le gène de référence et les télomères à partir d'un échantillon de référence. Pour mettre en place le protocole sur l'appareil LightCycler480 (Roche) dont nous disposons au laboratoire, je me suis appuyée sur le protocole publié par Cawthon et coll.²⁶⁹. J'ai utilisé l'échantillon d'ADN d'un individu non porteur d'une mutation dans le gène *ATM* et indemne de cancer pour établir la gamme d'étalonnage et tester les amorces de q-PCR pour l'amplification des télomères et du gène codant l'albumine, utilisé comme gène de référence (Annexe 9). J'ai réalisé la gamme étalon en utilisant une dilution séquentielle de l'ADN témoin (150ng, 50ng, 16.7ng, 5.55ng 1.87ng et 0ng). L'amplification par q-PCR des télomères et du gène de référence a été faite en triplicats en appliquant les recommandations de Promega pour le mélange réactionnel « GoTaq + qPCR Master Mix » (Annexe 9).

Afin de vérifier la qualité de la gamme étalon et la sensibilité de la technique, j'ai utilisé comme contrôle interne l'ADN des 5 lignées cellulaires suivantes dont la longueur des télomères avait été déterminée par l'équipe du Dr Arturo Londoño-Vallejo (Télomères et Cancer, Institut Curie) par la méthode de Southern blot : lignée 1080 (5kb), lignée HCA2 (8kb), lignée HeLa (14 à 31kb), lignée 1080ST (40kb) et lignée 1301 (80kb).

Les résultats d'optimisation du protocole (Figure 27) obtenus montrent que les analyses sont reproductibles car les coefficients directeurs des droites étalons varient très peu pour le gène de référence ($r^2=0,9966$, $r^2=0,9992$, $r^2=0,9979$) et pour les télomères ($r^2=0,9942$, $r^2=0,9934$, $r^2=0,9938$). Par ailleurs, les résultats obtenus pour les lignées cellulaires montrent que le seuil de détection de la fluorescence varie peu pour le gène de référence ($Ct_{\text{lignée 1301}}=27,5$, $Ct_{\text{lignée 1080ST}}=28,8$, $Ct_{\text{lignée HeLa}}=27,1$, $Ct_{\text{lignée HCA}}=27,6$, $Ct_{\text{lignée 1080}}=28,47$) alors que pour les télomères on observe des différences selon que la lignée cellulaire a des télomères longs ($Ct_{\text{lignée 1301}}=13,49$) ou des télomères courts ($Ct_{\text{lignée HCA2}}=18,86$).

L'optimisation de la technique de q-PCR sur le LightCycler480 montre que la q-PCR simplexe est une technique assez sensible qui génère des résultats reproductibles.

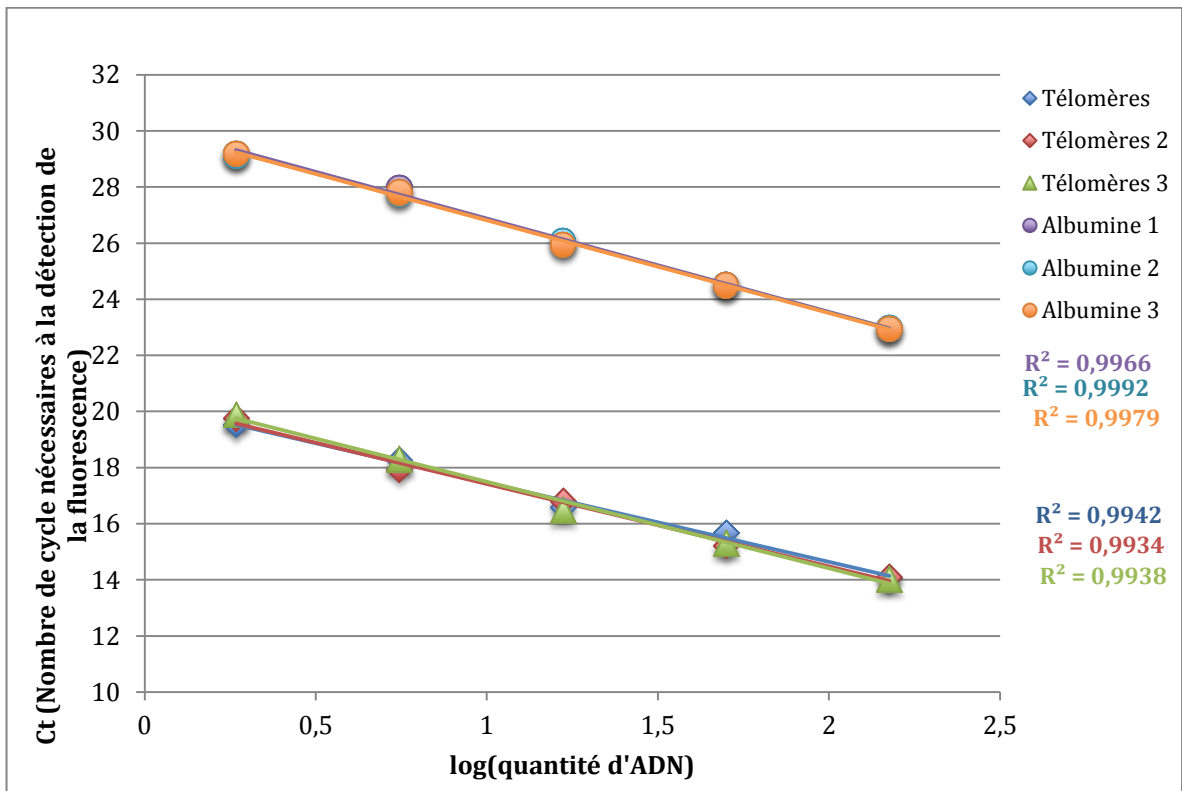


Figure 27 : Réalisation de la gamme étalon.

Chaque mesure a été réalisée en triplicat. Les coefficients directeurs (R^2) des droites confirment la linéarité et la reproductibilité des droites étalons.

4.7 Sélection des échantillons

Pour comparer la longueur relative des télomères des sujets HetAT et non-HetAT, j'ai sélectionné tous les participants des études CoF-AT et Retro-AT (décrites dans la partie I.II), hommes, femmes et enfants atteints d'A-T, ainsi que les individus vus en conseil génétique, pour lesquels un échantillon d'ADN était disponible. Afin d'éviter les biais dus au traitement par radiothérapie ou chimiothérapie j'ai exclu 31 participants ayant développé un cancer avant le prélèvement sanguin, soit 6 participants non-HetAT et 23 participants HetAT (Tableau 1 du manuscrit p.107).

Au total, j'ai mesuré la longueur des télomères de 427 participants dont 396 individus indemnes de cancer (236 individus HetAT et 160 individus non-HetAT) et 31 participants ayant développé un cancer après le prélèvement sanguin (25 individus HetAT et 6 non-HetAT). Les détails sur la répartition des participants en fonction du statut mutationnel d'*ATM*, de l'âge au prélèvement et de la localisation des cancers sont précisés dans le Tableau 1 de l'Article 1 rapportant sur les résultats de cette analyse.

5. Rôle modificateurs des SNPs associés au risque de cancer du sein et/ou à la longueur des télomères dans la survenue d'un cancer chez les femmes de la cohorte CoF-AT ou ayant participées à l'étude Rétro-AT

5.1 SNPs et longueur des télomères

Il y a 10 ans, Savage *et al* émettaient déjà l'hypothèse que le risque de cancer du sein pouvait être modulé par des SNPs localisés dans les gènes de la maintenance des télomères. Leur étude menée sur 1995 cas et 2296 témoins et s'intéressant à 31 SNPs avec une FAM supérieure à 5% dans la population générale concluait qu'il n'y avait pas de corrélation entre ces SNPs et le risque de cancer du sein²⁷³. Les limites de cette première étude étaient que les patientes étaient des cas rétrospectifs, prélevés après le traitement du cancer, et que différentes techniques avaient été utilisées pour le génotypage des tous les individus.

Il faut attendre l'avènement des études pan-génomiques (ou GWAS pour *Genome-Wide Association Studies*) dans les années 2010 pour mettre en évidence des associations entre des SNPs et la longueur des télomères. Codd *et coll.* notamment, ont réalisé un GWAS chez 37 684 hommes et femmes européens, sans antécédents médicaux connus. Dans cette étude, la longueur moyenne des télomères avait été mesurée à partir d'ADN sanguin. Les auteurs ont mis en évidence une association significative entre 7 SNPs et des télomères courts : rs10936599 (*TERC*), rs2736100 (*TERT*), rs7675998 (*NAF1*), rs9420907 (*OBFC1*), rs8105767 (*ZNF208*), rs755017 (*RTEL1*) et rs11125529 (*ACYP2*)²⁷⁴. Cette étude montre que les effets par allèle sur la diminution de la longueur des télomères sont faibles et varient entre 2 et 4pb par an.

En parallèle de la mise en évidence de SNPs associés à la longueur des télomères, Bojesen *et al* utilisaient les données de la puce à ADN iCOGS¹²² pour génotyper 480 SNPs localisés au locus *TERT* chez 53 752 témoins, 103 991 cas de cancer du sein provenant des études du BCAC (*Breast Cancer Association Consortium*) et de CIMBA (*Consortium of Investigators of Modifiers of BRCA1/2*), 39 774 cas de cancer de l'ovaire provenant d'OCAC (*Ovarian Cancer Association Consortium*) et 11 705 porteurs d'une mutation dans le gène *BRCA1*. Cette étude met en évidence deux SNPs associés aux télomères longs (rs7705526 et rs7726159) mais également une association entre le SNP rs7705526 et le risque de carcinome séreux de l'ovaire de bas grade¹²³. Une autre étude exploitant les données de la puce iCOGS a permis de mettre en évidence une association entre télomères courts et les SNPs rs1317082 (*TERC*), rs7726159 (*TERT*), rs6772228 (*PXK*), rs9257445 (*ZNF311*) et rs6060627 (*BCL2L1*) ainsi qu'une association entre le SNP rs2487999 (*OBFC1*) et les télomères longs²⁷⁵.

5.2 SNPs associés au risque de cancer du sein ou de l'ovaire

L'analyse des SNPs localisés au locus *TERT* a montré qu'en plus de l'association avec la longueur des télomères certains SNPs sont associés au risque de cancer du sein (rs3215401) ou au risque de cancer de l'ovaire (rs10069690) de façon générale, mais qu'ils modulent également le risque associé à ces cancers chez les porteurs *BRCA1* (rs10069690, rs2242652, rs2736108). D'autres SNPs également localisés au locus *TERT* sont associés avec le statut des récepteurs aux œstrogènes au niveau des tumeurs du sein, comme par exemple les SNPs rs2242652 et rs2736107 associés respectivement aux tumeurs RO- et RO+ ¹²³.

Comme nous l'avons vu dans l'introduction, plusieurs études, dont celles du BCAC, ont montré à ce jour une association entre une centaine de SNPs et le risque de cancer du sein dans la population générale. Par ailleurs, certains de ces SNPs modulent le risque de cancer du sein ou de l'ovaire chez les porteurs de mutations *BRCA1* ou *BRCA2*, mais sont également associés à des caractéristiques tumorales spécifiques, augmentant le risque de développer soit des tumeurs œstrogènes-positives (rs10941679^{276, 277}, rs6504950²⁷⁸, rs13281615^{278, 279}...), soit des tumeurs triple-négatives (rs13387042²⁷⁸, rs3803662²⁸⁰...) ou bien un cancer du sein de type lobulaire invasif (rs10995190)²⁸¹

5.3 SNPs d'intérêt dans les familles A-T et génotypage

5.3.1 Sélections des SNPs d'intérêts

L'ensemble des données rapportées sur les SNPs montre donc qu'au delà de leur association avec la longueur des télomères ou le cancer du sein, certains de ces variants peuvent également modifier le risque de cancer du sein ou de l'ovaire chez des porteurs de mutation dans un gène de prédisposition forte tel que *BRCA1* ou *BRCA2*. Nous avons donc émis l'hypothèse que des SNPs associés à la longueur des télomères ou au cancer du sein dans la population générale pouvaient également moduler le risque de cancer des porteurs hétérozygotes d'une mutation d'*ATM*. Etant donné la petite taille de notre échantillon d'étude (284 femmes HetAT et 174 femmes non HetAT), nous nous sommes cantonnés à l'étude de SNPs candidats relativement fréquents dans la population générale d'origine caucasienne (FAM \geq 1%).

Après avoir effectué une revue exhaustive de la littérature je me suis intéressée plus particulièrement à 25 SNPs dont l'implication dans la prédisposition au cancer du sein ou dans la maintenance de la longueur des télomères, a été montrée dans des études pan-génomiques menées notamment par les consortia BCAC et CIMBA incluant des milliers de sujets et/ou dont l'association a été répliquée dans plusieurs publications. Sept SNPs associés à la longueur des télomères dans la population générale (Tableau 12) ainsi que 15 autres SNPs associés au risque de cancer du sein dans

des études cas/témoins (Tableau 13) ont ainsi été sélectionnés.

5.3.2 Génotypage

J'ai effectué le génotypage avec la méthode Taqman qui permet de déterminer le génotype d'un individu à partir de la détection et la quantification de marqueurs fluorescents. Les sondes ciblant le polymorphisme sont composées d'un rapporteur fluorescent en 5' et d'un stabilisateur non-fluorescent à l'extrémité 3'. Une courte séquence du génome contenant le SNP d'intérêt est amplifiée par PCR en présence des sondes spécifiques aux deux allèles. Lors de cette amplification les sondes sont dégradées par la Taq polymérase, ce qui libère le rapporteur fluorescent. Au terme de la PCR l'intensité de fluorescence est mesurée par excitation à la lumière des agents fluorescents afin de définir les génotypes de chaque échantillon. Le protocole du génotypage par la méthode TaqMan est détaillée dans l'Annexe 10.

Pour la petite anecdote, la méthode TaqMan doit son nom au célèbre jeu vidéo PacMan (Taq polymérase + PacMan) car la Taq polymérase avance progressivement sur l'ADN et dégrade la sonde fluorescente de la même façon que la petite bouche jaune du PacMan sur les fruits.

Tableau 12 : SNPs candidats associés à la longueur des télomères dans d'autres études et sélectionnés.

SNPs	Locus	Gène associé	Changement allélique	FAM 1000 Genomes	Autres associations rapportées
rs2736100 ^{274, 282}	5p15.33	<i>TERT</i>	C>A	0.50	Cancer du poumon ²⁸³ Gliome ²⁸⁴
rs1317082 ²⁷⁵	3q26.2	<i>TERC</i>	A>G	0,24	-
rs2738783 ^{274, 275}	20q13.3	<i>RTEL1</i>	T>G	0,20	-
rs10165485 ^{274, 275}	2p16	<i>ACYP2</i>	T>C	0,13	-
rs2320615 ^{274, 275}	4q32.2	<i>NAF1</i>	G>A	0,26	-
rs9257445 ²⁷⁵	6p22.1	<i>ZNF311</i>	G>C	0,24	-
rs6060627 ²⁷⁵	18q21.3	<i>BCL2</i>	C>T	0,18	-

FAM= fréquence de l'allèle mineur

Tableau 13 : SNPs candidats associés au risque de cancer du sein dans d'autres études et sélectionnés.

SNPs	Locus	Gène associé	Changement allélique	FAM 1000 Genomes	Autres associations rapportées
rs11249433 ²⁸⁵	1q11.2	<i>EMBP1</i>	A>G	0,40	Cancer du sein chez les porteurs <i>BRCA2</i> ¹²⁶ Tumeurs RO+ ²⁸⁶
rs4973768 ¹²⁹	3p22	<i>SLC4A7</i>	C>T	0,50	Cancer du sein chez les porteuses <i>BRCA2</i> ²⁸⁷ Tumeurs RO+ ²⁷⁸
rs10941679 ²⁸⁸	5p12	<i>FGF10/MRPS30</i>	A>G	0,28	Tumeurs RO+ ^{276, 277}
rs2046210 ²⁸⁹	6q25.1	<i>ESR1</i>	G>A	0,43	Tumeurs RO- des porteurs <i>BRCA1/2</i> ²⁹⁰
rs1011970 ¹²⁷	9p21	<i>CDKN2A</i>	G>T	0,17	Cancer du sein <i>in situ</i> ²⁹¹ Cancer du sein chez l'homme ²⁹²
rs2380205 ²⁹³	10p15	<i>ANKRD16</i>	C>T	0,36	-
rs10995190 ¹²⁷	10q21	<i>ZNF365</i>	G>A	0,17	Cancer du sein chez les porteurs <i>BRCA2</i> ¹²⁷ Cancer du sein lobulaire ²⁸¹
rs704010 ¹²⁷	10q22	<i>ZMIZ1</i>	C>T	0,36	-
rs13387042 ^{288, 294}	2q35	<i>IGFPB5</i>	G>A	0,45	Porteur <i>BRCA1/2</i> ²⁹⁵ Tumeurs triples-négatives ²⁷⁸ Tumeurs basales ²⁷⁸
rs13281615 ²⁹⁶	8q24	<i>CASC8</i>	A>G	0,44	Tumeurs RO+ ^{278, 279} Tumeurs de bas grade ²⁷⁹
rs2981578 ²⁹⁷	10q26	<i>FGFR2</i>	C>T	0,51	-
rs1045485 ¹²¹	2q33-q34	<i>CASP8</i>	G>C	0,11	Tumeurs PR- ²⁷⁸
rs889312 ²⁹⁷	5q11.2	<i>MAP3K1</i>	A>C	0,33	Porteurs <i>BRCA2</i> ¹²⁸ Cancer lobulaire ²⁸¹ Tumeurs RO- ²⁷⁹
rs3803662 ²⁹⁷	16q12.1	<i>TNRC9</i>	G>A	0,29	Porteurs <i>BRCA1/2</i> ¹²⁸ Tumeurs triples-négatives ²⁸⁰ Cancer du sein chez l'homme ²⁹⁸
rs3817198 ²⁹⁷	11p15.5	<i>LSP1</i>	T>C	0,32	Porteurs <i>BRCA2</i> ¹²⁹ Tumeurs triples-négatives ²⁷⁸

FAM : Fréquence de l'Allèle Mineur

5.4 Analyse statistique

Les associations entre SNPs et risque de cancer, ou SNPs et longueur des télomères ont été testées par régressions logistiques en ajustant le cas échéant, pour différents facteurs : le genre, l'âge au diagnostic pour les cas de cancers et l'âge aux dernières nouvelles pour les participants indemnes.

Pour les SNPs, un modèle co-dominant (ou additif) a été examiné pour estimer les odds ratios (OR) et obtenir la valeur de p (« p-value »). Ce modèle consiste à tester l'association du nombre de copies de

l'allèle à risque (0, 1, 2) avec le phénotype étudié ; ceci diffère du modèle dominant (1 ou 2 copies versus 0 copie) et du modèle récessif (2 copies versus 0 ou 1 copies). J'ai utilisé le logiciel STATA (version 14.2, StataCorp) pour réaliser toutes les analyses.

6. ARTICLE 1 : résultats de l'analyse sur le rôle de la longueur des télomères et des SNPs associés à la longueur des télomères sur le risque de cancer chez les sujets HetAT.

ORIGINAL ARTICLE

Telomere length, ATM mutation status and cancer risk in Ataxia-Telangiectasia families

Anne-Laure Renault^{1,2,3,4}, Noura Mebirouk^{1,2,3,4}, Eve Cavaciuti^{1,2,3,4},
Dorothee Le Gal^{1,2,3,4}, Julie Lecarpentier⁵, Catherine Dubois d'Enghien⁶,
Anthony Laugé⁶, Marie-Gabrielle Dondon^{1,2,3,4}, Martine Labbé^{1,2,3,4}, Gaetan Lesca⁷,
Dominique Leroux⁸, Laurence Gladieff⁹, Claude Adenis¹⁰, Laurence Faivre¹¹,
Brigitte Gilbert-Dussardier¹², Alain Lortholary¹³, Jean-Pierre Fricker¹⁴, Karin Dahan¹⁵,
Jacques-Olivier Bay¹⁶, Michel Longy¹⁷, Bruno Buecher⁶, Nicolas Janin¹⁵,
Hélène Zattara¹⁸, Pascaline Berthet¹⁹, Audrey Combès²⁰, Isabelle Coupier^{21,22}, CoF-AT
study collaborators^{1,2,3,4}, Janet Hall^{23,24,25}, Dominique Stoppa-Lyonnet^{6,26,27},
Nadine Andrieu^{1,2,3,4} and Fabienne Lesueur^{1,2,3,4,*}

¹INSERM, U900, Paris, France, ²PSL Research University, Paris, France, ³Institut Curie, Paris, France, ⁴Mines Paris Tech, Fontainebleau, France, ⁵Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, ⁶Service de Génétique, Institut Curie, Paris, France, ⁷CHU de Lyon, Groupement Hospitalier Est, Service de Génétique Médicale, Lyon, France, ⁸CHU de Grenoble, Hôpital Couple-Enfant, Département de Génétique, Grenoble, France, ⁹Institut Claudius Regaud-IUCT-Oncopole, Service d'Oncologie Médicale, Toulouse, France, ¹⁰Centre Oscar Lambret, Lille, France, ¹¹Hôpital d'Enfants, Service de Génétique Médicale, Dijon, France, ¹²CHU de Poitiers, Service de Génétique Médicale, Poitiers, France, ¹³Centre Catherine de Sienne, Service d'Oncologie Médicale, Nantes, France, ¹⁴Centre Paul Strauss, Unité d'Oncologie, Strasbourg, France, ¹⁵Clinique Universitaire Saint-Luc, Génétique, Bruxelles, Belgium, ¹⁶CHU Clermont-Ferrand site Estaing, Clermont-Ferrand, France, ¹⁷Institut Bergonié, Bordeaux, France, ¹⁸Hôpital Saint-Joseph, Marseille, France, ¹⁹Centre François Baclesse, Unité de Pathologie Gynécologique, Caen, France, ²⁰Centre Hospitalier Universitaire de Nîmes, Unité de Génétique Médicale et Cytogénétique, Nîmes, France, ²¹Hôpital Arnaud de Villeneuve, CHU Montpellier, Service de Génétique Médicale et Oncogénétique, Montpellier, France, ²²ICM Val d'Aurel, Unité d'Oncogénétique, Montpellier, France, ²³Centre de Recherche en Cancérologie de Lyon, Lyon, France, ²⁴UMR INSERM 1052, Lyon, France, ²⁵CNRS 5286, Lyon, France, ²⁶INSERM, U830, Paris, France and ²⁷Université Paris Descartes, Paris, France.

*To whom correspondence should be addressed. Tel: +33 1 72 38 93 86; Fax: +33 1 53 10 26 63; Email: fabienne.lesueur@curie.fr

Abstract

Recent studies have linked constitutive telomere length (TL) to aging-related diseases including cancer at different sites. ATM participates in the signaling of telomere erosion, and inherited mutations in ATM have been associated with increased risk of cancer, particularly breast cancer. The goal of this study was to investigate whether carriage of an ATM mutation and TL interplay to modify cancer risk in ataxia-telangiectasia (A-T) families.

The study population consisted of 284 heterozygous ATM mutation carriers (HetAT) and 174 non-carriers (non-HetAT) from 103 A-T families. Forty-eight HetAT and 14 non-HetAT individuals had cancer, among them 25 HetAT and 6 non-HetAT were diagnosed after blood sample collection. We measured mean TL using a quantitative PCR assay and genotyped seven single-nucleotide polymorphisms (SNPs) recurrently associated with TL in large population-based studies.

Received: January 8, 2017; Revised: June 21, 2017; Accepted: July 8, 2017

© The Author 2017. Published by Oxford University Press.

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

HetAT individuals were at increased risk of cancer (OR = 2.3, 95%CI = 1.2–4.4, $P = 0.01$), and particularly of breast cancer for women (OR = 2.9, 95%CI = 1.2–7.1, $P = 0.02$), in comparison to their non-HetAT relatives. HetAT individuals had longer telomeres than non-HetAT individuals ($P = 0.0008$) but TL was not associated with cancer risk, and no significant interaction was observed between ATM mutation status and TL. Furthermore, rs9257445 (ZNF311) was associated with TL in HetAT subjects and rs6060627 (BCL2L1) modified cancer risk in HetAT and non-HetAT women.

Our findings suggest that carriage of an ATM mutation impacts on the age-related TL shortening and that TL *per se* is not related to cancer risk in ATM carriers. TL measurement alone is not a good marker for predicting cancer risk in A-T families.

Abbreviations

ALB	Albumin
ATM	Ataxia-Telangiectasia Mutated
A-T	Ataxia-Telangiectasia
BC	breast cancer
REF	Reference
RTL	relative telomere length
TEL	telomere repeat units
TL	telomere length

Introduction

Biallelic mutations in the *Ataxia-Telangiectasia Mutated* (ATM) gene are responsible for Ataxia-Telangiectasia (A-T), a rare autosomal recessive disorder characterized by progressive neuronal degeneration, immunological deficiency, genetic instability, radiosensitivity and an increased risk of cancer. Epidemiological studies on A-T families showed that heterozygous mutation carriers (hereafter called ‘HetAT’ subjects) are at increased risk of cancer (1), particularly of breast cancer (BC) in female relatives (2–4). Moreover, recent studies conducted in familial or early onset BC cases and ethnically similar unrelated controls have confirmed ATM as a BC susceptibility gene in a context independent of a genetic predisposition to A-T (5,6).

The ATM gene encodes a protein kinase, which is a central regulator of the cellular response to DNA damage and also the response to telomere dysfunction. Telomeres are repeats units of TTAGGG nucleotides coated by a six-subunit protein complex named shelterin that comprises TRF1, TRF2, POT1, TPP1, TIN2 and RAP1. The telomeric complex caps the ends of chromosomes to maintain genome integrity by protecting chromosomes from end-to-end fusion and exonuclease degradation (7,8). In normal cells, telomere elongation is telomerase-dependent and is highly regulated by various proteins including ATM which is a positive regulator of telomere lengthening (9,10). The phosphorylation of the shelterin protein TRF1 by ATM diminishes the interaction between TRF1 and telomeric DNA (11) leading to the release of TRF1 and ensuring access for the telomerase enzyme to lengthen telomeres (12). More recently, it was shown that the shelterin complex provides protection of chromosome ends against classical nonhomologous end-joining (NHEJ) and homologous-direct repair (HDR) by repressing the ATM and ATR pathway signaling (13).

Telomeres have long been recognized as a biomarker of cell aging due to their progressive shortening with each cell division (14,15), and a number of studies have demonstrated clear involvement of telomere shortening in aging-related diseases, in particular in the carcinogenesis of several malignancies (16). However, the direction of association between telomere length (TL) and cancer risk appears to be cancer-site dependent. For instance, longer telomeres have been associated with

increased risk of gastric (17), lung (18), bladder (19) and esophageal cancers (20), while shorter telomeres have been associated with increased risk of non-Hodgkin lymphoma (21) and melanoma (22). The association between constitutive TL and BC risk remains controversial: some studies have associated shorter telomeres with an increased risk of BC (23,24), other studies have associated this risk with longer telomeres (25,26), while the prospective Sister Study did not support TL as a biomarker for BC risk (27). Results from the EMBRACE study confirmed the null association between TL and breast or ovarian cancer risk in BRCA1 and BRCA2 mutation carriers although longer lymphocyte TL was observed in this high-risk population than in the non-carriers population (28). Furthermore, several studies conducted in the general population have shown a significant association between TL and SNPs located at different loci (29,30), including rs2736100 in TERT (29), rs1317082 in TERC (30), rs2738783 in RTEL1 (30), rs10165485 in ACYP2 (30), rs2320615 in NAF1 (30), rs9257445 in ZNF311 (30) and rs6060727 in BCL2L1 (30). Some of the TL-related SNPs have also been associated with an increased risk of cancer (26,30,31).

Association between the ATM function and TL has been suggested by cytogenetic studies where lymphocytes with biallelic inactivation of ATM were characterized by an abnormal rate of chromosome end-to-end fusions (32,33) suggesting that the disruption of the telomere lengthening might be the cause of genome instability and disorder progression observed in lymphocytes from A-T patients (12). However to our knowledge, no study has examined lymphocyte TL in HetAT subjects. The goal of the present study was to assess whether being a heterozygous carrier of an ATM mutation influences TL, and to examine how ATM mutation status and TL interplay to modify cancer risk in HetAT subjects.

Material and methods

Study participants

Study participants were all related to a child affected with ataxia-telangiectasia (A-T) and were enrolled through one of the three following French studies: the Prospective Cohort of Women Related to an A-T child (CoF-AT) (Lecarpentier J and Andrieu N, *personal communication*), the Retrospective Study on A-T families (Retro-AT) (34), or the A-T families collection assembled by the Genetic Department of Institut Curie, which is the national reference center for genetic diagnosis of A-T.

Retro-AT was carried out between 1994 and 1997 to assess cancer risk in A-T families living in France (3,34,35). Thirty-four A-T families were identified during this period and 27 of them were subsequently included in the CoF-AT prospective study. CoF-AT is an ongoing national prospective cohort, which was initiated in 2003 to investigate environmental and genetic risk factors for breast cancer (BC) in HetAT and non-HetAT women. All women aged 18 and over are eligible to participate in the study. At inclusion, participants provide a blood sample to determine whether they carry the ATM mutations identified in the affected child in that family. Women are followed up for 10 years after inclusion, and re-contacted every 2 years to provide information on medical data, environmental and lifestyle factors

through a detailed epidemiological questionnaire. Three hundred and forty six women (177 HetAT and 169 non-HetAT) belonging to 96 A-T families from CoF-AT were included in the present study. Among these, three CoF-AT families were known to segregate a mutation in the BRCA1 or BRCA2 gene in addition to the familial ATM mutation. Carriers of a mutation in one of these two major BC predisposing genes were excluded from the subsequent analyses (three HetAT and one non-HetAT women). Sixty-seven additional subjects related to an A-T child and who attended the family genetics clinics for molecular diagnosis were also included. Written informed consent was obtained from all participants. The Retro-AT and CoF-AT study protocols were approved by the appropriate ethics committee (CCP Ile-de-France III) and by the French data protection authority (CNIL).

In total, the study was conducted on 458 individuals (284 HetAT and 174 non-HetAT) with available blood DNA sample. Among those, 62 subjects (48 HetAT and 14 non-HetAT) had cancer, and 31 of them were diagnosed after blood sample was collected (incident cases: 25 HetAT and 6 non-HetAT) (Table 1).

DNA preparation and characterization of ATM mutation status

Genomic DNA was extracted from whole blood samples collected from EDTA with the Quickgene 610-L automated system (Fujifilm) according to the manufacturer's instructions or, for less than 10% of the study participants, using other silica-based DNA purification kits (Macherey Nagel or QIAGEN kits) or standard phenol-chloroform procedure. For all participants the presence of the familial ATM mutation identified in the A-T child was determined according to their parental branch by targeted sequencing on an ABI3500xL DNA analyzer (Applied Biosystems).

Selection of subjects for SNP genotyping and telomere length measurement

While all 458 participants were included in the SNP analysis, TL analysis was restricted to the 427 participants (261 HetAT and 166 non-HetAT) who were unaffected with cancer when biological material was collected (Table 1).

Telomere length measurement

The relative telomere length (RTL) was ascertained by quantitative PCR on the LightCycler480 (Roche) using the protocol developed by Cawthon et al. (36). The RTL was determined as the ratio of detected fluorescence from the amplification of telomere repeat units (TEL) relative to that of a single-copy reference sequence from the housekeeping gene ALB encoding Albumin (REF).

Telomere repeat units and the ALB amplicon were amplified in separate wells on the same 384-well plate to avoid the inter-plate variability. A five-point standard curve was established for each plate with one reference genomic DNA sample diluted in 3-fold series (150 ng, 50 ng, 16.7 ng, 5.55 ng and 1.85 ng per well). The standard curves were then used to determine the TEL and REF amplified DNA amount and corresponding dilution factors, which were used to calculate the TEL/REF ratio representing the RTL. Telomeres and ALB amplifications were performed in duplicate for each individual's DNA sample. Two blind replicate samples were interspersed with the samples on each plate to assess the inter-plate variability. Samples with failed amplification of telomere repeat units and/or ALB were not repeated. The PCR efficiency for both telomeres and ALB amplifications was greater than 97%.

Table 1. Distribution of gender, age, cancer cases and cancer type in the 103 A-T families, by ATM mutation status and sample series

	CoF-AT (N = 346)*		Retro-AT (N = 45)		IC Collection (N = 67)	
	Non-HetAT (N = 169)	HetAT (N = 177)	Non-HetAT (N = 1)	HetAT (N = 44)	Non-HetAT (N = 4)	HetAT (N = 63)
Gender						
Female	169 ^a	177 ^{a,b}	1	15	1	11
Male	0	0	0	29	3	52
Mean age at blood draw (se)	42.3 (1.2)	42.7 (1.0)	35.3 (N/A)	42.5 (1.5)	46.1 (8.6)	43.7 (1.6)
Age at blood draw (years)						
≤50	120	126	1	37	2	48
>50	49	51	0	7	2	15
Cancer status						
Unaffected females	155	149 ^b	1	12	1	7
Affected females	14 (6) ^a	28 (10) ^{a,b}	0	3 (2)	0	4 (3)
Unaffected males	0	0	0	22	3	46
Affected males	0	0	0	7 (5)	0	6 (5)
Cancer Type						
Breast	8 (3) ^a	18 (7) ^{a,b}	—	2 (1)	—	4 (3)
Cervix	—	2 (1)	—	—	—	—
Colon-rectum	—	1 (0)	—	—	—	—
Genital organs	—	1 (0)	—	—	—	—
Kidney	2 (2)	1 (0)	—	—	—	—
Leukemia	—	—	—	—	—	1 (1)
Lung	1 (0)	—	—	1 (0)	—	1 (0)
Melanoma	—	1 (0)	—	—	—	1 (1)
Non-Hodgkin lymphoma	1 (0)	1 (0)	—	1 (1)	—	—
Ovarian	1 (0)	—	—	—	—	—
Pancreas	—	1 (1)	—	—	—	1 (1)
Prostate	—	—	—	5 (4)	—	2 (2)
Stomach	—	—	—	1 (1)	—	—
Thyroid	—	2 (1)	—	—	—	—
Uterus	1 (1)	—	—	—	—	—

The number of cancer cases eligible for the telomere length analysis is indicated in brackets and in italics.

se: Standard error.

^aOne woman from this group carried a BRCA1 mutation and was excluded in the subsequent analyses.

^bOne woman from this group carried a BRCA2 mutation and was excluded in the subsequent analyses.

SNP genotyping

The SNPs rs2736100 (TERT) (29), rs1317082 (TERC) (30), rs2738783 (RTEL1) (30), rs10165485 (ACYP2) (30), rs2320615 (NAF1) (30), rs9257445 (ZNF311) (30) and rs6060727 (BCL2L1) (30) which have been recurrently associated with TL in the literature were selected. These seven SNPs were genotyped for all participants using a standard Taqman protocol and the fluorescence was read on an ABI PRISM7900HT instrument (Applied Biosystems). Taqman probes targeting alleles of the seven SNPs are available from the ThermoFisher Scientific website (<http://www.thermoFisher.com/order/genome-database/>). Ten percent of the samples were randomly duplicated between 384-well plates to validate the results reproducibility. The proportion of successfully genotyped DNA samples was above 95% for all SNPs except for rs2736100 for which the call rate was 90%.

Statistical analysis

Variance analyses (ANOVA) were performed to estimate the intra- and inter-plates assay variability. Linear regressions were used to evaluate the correlation between RTL and age at blood sampling for HetAT and non-HetAT subjects, and a likelihood ratio test was performed to determine if the regression coefficients were significantly different between the two groups.

Logistic regressions were used to obtain the odds ratios (OR) and 95% confidence intervals (CIs) for the association between cancer risk and RTL, as well as for the association between cancer risk and ATM mutation status.

In the SNP analysis, the deviation of the genotypes proportion from Hardy-Weinberg Equilibrium (HWE) was assessed in the unaffected subjects using chi-square test with one degree of freedom and no tested SNPs showed significant HWE deviation with $P < 0.05$. We used unconditional logistic regression models to estimate ORs and the trend p-value for association between cancer risk and each TL-related SNP, using codominant coding for genotypes (0,1,2) with the homozygote of the common allele as the reference group. The associations between TL and SNPs were assessed using a linear regression with dominant coding for genotypes (no effect allele = 0, at least one effect allele = 1) and grouping subjects with medium and short telomeres together.

All analyses were adjusted on gender (when relevant), age at last interview or age at diagnosis, whichever came first. All statistical tests were two-sided and data was analyzed using STATA, version 14.1 (StataCorp, College Station, TX).

Results

ATM mutation status and cancer risk

Three CoF-AT families were known to segregate a BRCA1 or BRCA2 mutation in addition to the familial ATM mutation. After excluding BRCA1 or BRCA2 mutation carriers, 281 HetAT and 173 non-HetAT participants were included in the genetic study. We confirmed in this study set that carriers of a mutated copy of ATM had a higher overall cancer risk (OR = 2.3, 95% CI = 1.2–4.4, $P = 0.01$). In particular, women had a 3-fold increase in risk of BC, the most frequent cancer in our study sample (OR = 2.9, 95% CI = 1.2–7.1, $P = 0.02$) (Table 2). Even though most of the subjects

were enrolled prospectively in the study, 22 of the 59 cancer cases were prevalent cases. As a result, the observed association could reflect a measure of susceptibility to both cancer onset and survival. We therefore also performed the analysis on incident cases only, and found similar results (OR = 3.3, 95% CI = 1.3–8.1, $P = 0.01$ and OR = 3.7, 95% CI = 1.0–13.5, $P = 0.04$, respectively) (Table 2).

TL and ATM mutation status

No intra-plate variability was observed within telomere assay duplicates and within ALB assay duplicates ($P = 0.07$ and $P = 0.05$, respectively), nor was inter-plate variability observed for the telomere assays or the ALB assays ($P = 0.24$ and $P = 0.20$, respectively) using an ANOVA test. Thus, subsequent analyses were performed without adjustment for plate location.

RTL was measured on all unaffected participants and on cases for whom lymphocyte DNA was obtained prior to the diagnosis of cancer (incident cases), as both cancer and treatment may affect telomere elongation. As expected, TL was negatively correlated with age at sampling in the control group of unaffected non-HetAT subjects ($R^2 = -0.0048$, $P = 0.003$). This tendency was also observed in unaffected HetAT subjects, although the rate of decline was not as steep as for their non-HetAT counterparts ($R^2 = -0.0015$, $P = 0.31$) (Figure 1). The difference between the slopes for TL with respect to age for HetAT and non-HetAT participants was significant when females and males were included in the analysis ($P = 0.0008$). Because it has been shown that females have longer telomeres than males (37,38), we next restricted the analysis to females. In this group TL showed a shortening with age but with a less-steep decline in HetAT than in non-HetAT females. However, no statistically significant difference was seen between the two groups ($P = 0.24$), which could be due to a power lowered by a smaller number of subjects (Figure 1). We could not conduct a similar comparison between unaffected HetAT and non-HetAT males because data for non-HetAT were too scarce. However, as observed in HetAT females, the shortening of the TL with age was not significant in unaffected HetAT males (data not shown). There were too few observations on HetAT and non-HetAT subjects affected with cancer for meaningful analyses to be carried out.

We also measured the RTL in a group of 73 A-T children belonging to the CoF-AT and Retro-AT families. As expected we found that they had longer TL than their HetAT and non-HetAT relatives due to their younger age (mean age = 10.6, standard error = 0.92) (data not shown). Therefore making further TL comparison between A-T children and the HetAT studied group was inadequate.

TL status and cancer risk

We first defined tertiles of RTL distribution among all unaffected individuals (Long > 1.2886; Medium: 1.1009–1.2886;

Table 2. Association between ATM mutation status and cancer risk, for all cancer sites combined and for breast cancer (BC) in females

Outcome	Subjects	All cancer cases					Incident cancer cases			
		Unaffected (N)	Affected (N)	OR ^a	(95% CI)	P	Affected (N)	OR ^a	(95% CI)	P
Overall cancer risk (all participants)	Non-HetAT	160	13	1			6	1		
	HetAT	235	46	2.3	(1.2, 4.4)	0.01	31	3.3	(1.3, 8.1)	0.01
Overall cancer risk (females only)	Non-HetAT	157	13	1			6	1		
	HetAT	167	33	2.4	(1.2, 4.7)	0.01	18	2.8	(1.1, 7.4)	0.03
BC risk (females only)	Non-HetAT	157	7	1			3	1		
	HetAT	167	22	2.9	(1.2, 7.1)	0.02	12	3.7	(1.0, 13.5)	0.04

^aOR adjusted for age at last interview/diagnosis, and for gender when relevant.

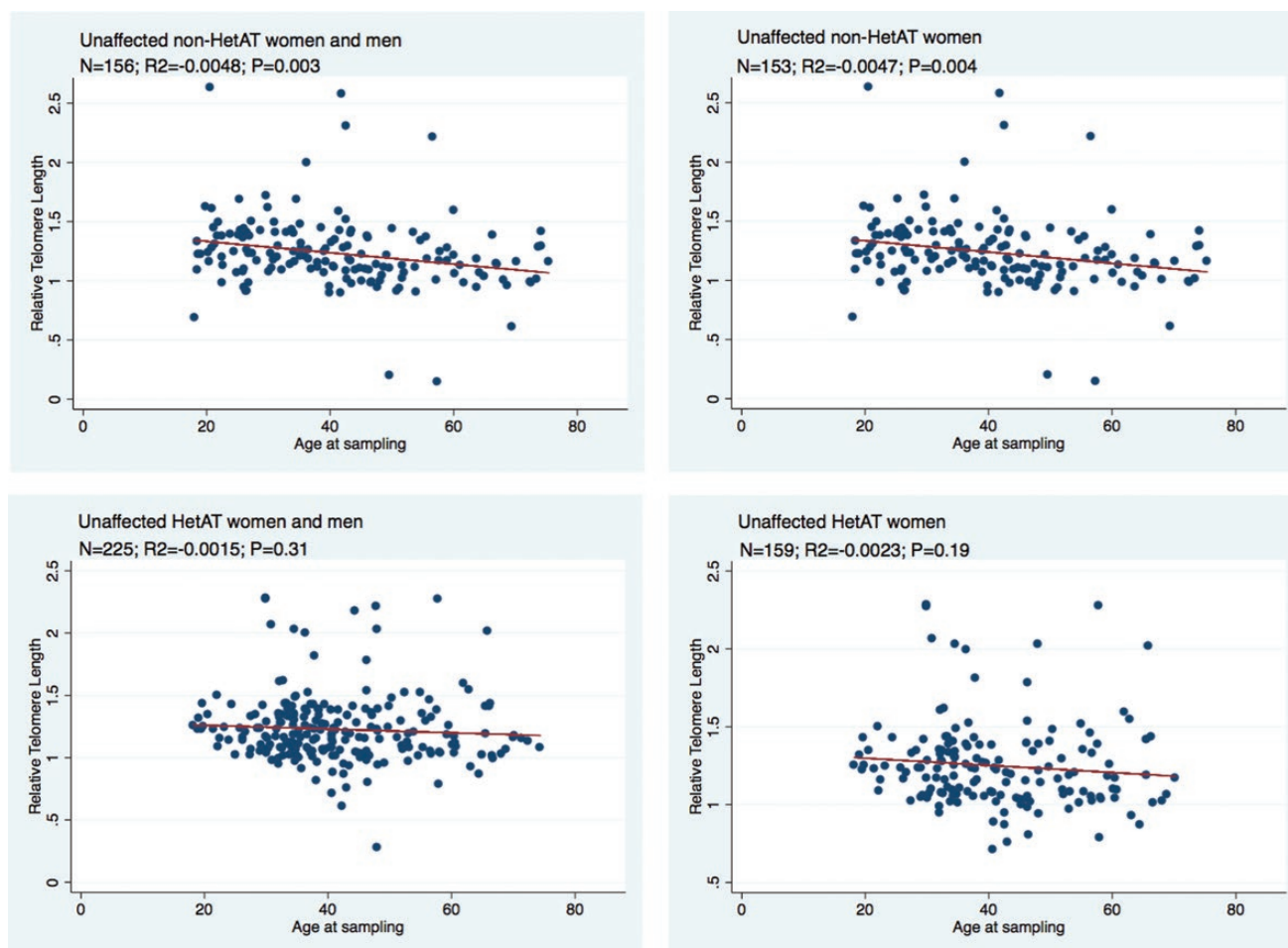


Figure 1. Correlations between relative telomere length and age at blood draw in unaffected HetAT and non-HetAT subjects. N, number of analyzed subjects. R^2 and P values (P) are for the correlation from the linear regression.

Short < 1.1009) and investigated the association between RTL and cancer occurrence using the longest telomere tertile as reference group. No significant association was found with overall cancer risk or with BC risk in women (Table 3). No interaction was evidenced between RTL and ATM mutation status for either overall cancer risk or BC risk (Table 3).

We next classified the subjects into four groups according to their ATM mutation status and RTL. Due to the limited number of observations in some groups, subjects with medium or short telomeres were pooled together; non-carriers with long telomeres were used as the reference group. Although results did not reach statistical significance, we found that HetAT subjects with shorter telomeres were more prone to develop cancer at any site than the other groups ($P_{\text{trend}} = 0.03$). The same trend was seen when restricting the analysis to BC risk ($P_{\text{trend}} = 0.05$) (Table 4).

Finally, we examined whether telomere-related SNPs were associated (i) with RTL in order to evaluate how those SNPs could serve as surrogate of RTL measurement in our population sample, and (ii) with cancer risk.

Since age-related shortening of TL was different between HetAT and non-HetAT subjects, we tested the association between SNPs and RTL in the two groups separately. Surprisingly, none of the tested SNPs showed association with RTL in the non-HetAT group (all the 162 non-HetAT subjects are females but three), while suggestive association between rs9257445 and RTL was observed in the HetAT group where both 162 females and

76 males were included in the analysis ($P = 0.03$) (Supplementary Table 1). In addition to rs9257445, two other SNPs, rs2320615 and rs10165485 showed significant association with RTL when analyses were conducted on the HetAT male subgroup ($P = 0.02$, $P = 0.04$ and $P = 0.02$, respectively).

Regarding the contribution of the seven selected SNPs and cancer risk, only rs6060627 in *BCL2L1* showed suggestive association with overall cancer risk, with a significant trend when the analysis was focused on females ($P_{\text{trend}} = 0.03$) (Table 5). No association between the seven SNPs and cancer risk was found when HetAT and non-HetAT subjects were analyzed separately (data not shown).

Discussion

Few studies have examined whether TL modified cancer risk in the context of a familial predisposition involving inherited mutations with incomplete penetrance. One study has assessed the association between cutaneous malignant melanoma and TL in melanoma-prone families with and without *CDKN2A* mutations. Authors found that longer telomeres increased the melanoma risk in non-carrier cases (39). Regarding familial breast cancer, Pooley et al. showed that carriers of a mutation in *BRCA1* or *BRCA2* have longer telomeres than their non-carrier relatives, but BC risk for carriers was not modified by the TL²⁸.

Table 3. Association between relative telomere length (RTL) and cancer risk, for all cancer sites, and for breast cancer in females

RTL tertile	Unaffected	Affected	OR	(95% CI)	P_{trend}	OR ^c	(95% CI)	P_{trend}^c	$P_{\text{interaction}}$
(all cancers)									
All participants	N = 381	N = 31							
1 st (long)	127	7	1 ^a			1 ^d			
2 nd (medium)	127	13	1.8	(0.7, 4.7)		1.9	(0.7, 4.9)		0.68
3 rd (short)	127	11	1.5	(0.5, 3.9)	0.48	1.5	(0.5, 3.9)	0.49	0.72
(all cancers)									
Females only	N = 312	N = 21							
1 st (long)	104	6	1 ^b			1 ^e			
2 nd (medium)	104	9	1.5	(0.5, 4.4)		1.68	(0.6, 5.0)		0.75
3 rd (short)	104	6	0.9	(0.3, 3.0)	0.91	0.96	(0.3, 3.1)	0.94	0.57
(breast cancer)									
Females only	N = 312	N = 14							
1 st (long)	104	4	1 ^b			1 ^e			
2 nd (medium)	104	6	1.5	(0.4, 5.5)		1.6	(0.4, 6.0)		0.93
3 rd (short)	104	4	1.0	(0.2, 4.1)	0.99	1.0	(0.2, 4.0)	0.97	0.99

^aAdjusted on gender and age at blood draw.^bAdjusted on age at blood draw.^cATM status was included in the model.^dAdjusted on gender, age at blood draw and ATM status.^eAdjusted on age at blood draw and ATM status.**Table 4.** Cancer risk according to ATM mutation status and RTL

Group	Unaffected	Affected	OR	(95% CI)	P	P_{trend}
(all cancers)						
All participants	N = 381	N = 31				
Non-HetAT + Long RTL	53 (13.9%)	1 (3.2%)	1 ^a			
Non-HetAT + Medium/Short RTL	103 (27.0%)	5 (16.1%)	2.5	(0.3, 22.3)	0.40	
HetAT + Long RTL	74 (19.4%)	6 (19.4%)	3.8	(0.4, 33.3)	0.22	
HetAT + Medium/Short RTL	151 (39.7%)	19 (61.3%)	5.8	(0.7, 45.5)	0.10	0.03 ^a
(all cancers)						
Females only	N = 312	N = 21				
Non-HetAT + Long RTL	53 (17.0%)	1 (4.8%)	1 ^b			
Non-HetAT + Medium/Short RTL	100 (32.0%)	5 (23.8%)	2.3	(0.3, 20.8)	0.44	
HetAT + Long RTL	56 (18.0%)	5 (23.8%)	4.2	(0.5, 37.7)	0.20	
HetAT + Medium/Short RTL	103 (33.0%)	10 (47.6%)	5.0	(0.6, 39.9)	0.13	0.06 ^b
(breast cancer)						
Females only	N = 312	N = 14				
Non-HetAT + Long RTL	53 (17.0%)	1 (7.2%)	1 ^b			
Non-HetAT + Medium/Short RTL	100 (32.0%)	2 (14.3%)	1.0	(0.1, 11.8)	0.97	
HetAT + Long RTL	56 (18.0%)	3 (21.4%)	2.8	(0.3, 27.7)	0.39	
HetAT + Medium/Short RTL	103 (33.0%)	8 (57.1%)	4.1	(0.5, 33.5)	0.19	0.05 ^b

^aAdjusted on age at last interview/diagnosis and gender adjustment.^bAdjusted on age at last interview/diagnosis.

To our knowledge, this is the first exploratory study investigating TL as a modifier of cancer risk for individuals carrying an ATM mutation. One asset of our study design is the homogeneity of the study sample, given that HetAT and non-HetAT participants were all related to an A-T child. Consequently only 'true' pathogenic ATM mutations, i.e. base substitutions, insertions or deletions that generate premature termination codons or splicing abnormalities were considered, thus avoiding variant misclassification. Moreover, most of the study participants were recruited prospectively, with blood samples collected prior to cancer diagnosis for 50% of the cases. Such a design is better able to address the question of whether telomeres become shorter before cancer diagnosis, and are therefore predictive of cancer development. Finally, all DNA samples were processed, stored and analyzed using the same workflow thus eliminating biases when measuring RTL.

Consistent with previous reports, and in particular with the French retrospective study conducted on 34 A-T families (3,34,35), we found in this larger family series that HetAT individuals had a 2- to 3-fold increased overall cancer risk as compared to non-HetAT individuals. Because BC was the most frequent cancer in the study population composed of 81.7% of women, we could only estimate risk for this specific cancer separately. We found that HetAT women had a 3-fold increase risk of developing BC and this was even higher when considering incident cases only, confirming previous findings in A-T families (3,4,35). About 1% of the general population and nearly 3% of hereditary breast and ovarian cancer patients are estimated to carry a deleterious ATM variant (5,6). It is therefore important to better estimate the risk of cancer associated with a mutation in this gene by taking into account other genetic and environmental/lifestyle factors

Table 5. Associations between telomere-related SNPs and cancer risk

	SNP	Genotype	Unaffected	Affected	OR	(95%CI)	P	P _{trend} ^c
All participants	rs2736100 (TERT)	CC	99	13	1 ^a			
		CA	155	24	1.18	(0.57, 2.46)	0.66	
		AA	91	10	0.9	(0.37, 2.17)	0.81	0.84
	rs1317082 (TERC)	AA	214	30	1 ^a			
		AG	130	18	0.87	(0.45, 1.66)	0.67	
		GG	18	6	2.39	(0.88, 6.53)	0.09	0.38
	rs2738783 (RTEL1)	TT	255	43	1 ^a			
		TG	102	11	0.67	(0.33, 1.35)	0.26	
		GG	12	2	1.03	(0.22, 4.79)	0.97	0.41
	rs10165485 (ACYP2)	TT	308	42	1 ^a			
		TC	52	12	1.78	(0.87, 3.62)	0.11	
		CC	7	1	1.09	(0.13, 9.13)	0.93	0.20
	rs2320615 (NAF1)	GG	203	33	1 ^a			
		GA	144	20	0.91	(0.50, 1.66)	0.75	
		AA	24	3	0.82	(0.23, 2.89)	0.76	0.68
	rs9257445 (ZNF311)	GG	206	36	1 ^a			
		GC	141	13	0.48	(0.24, 0.95)	0.035	
		CC	22	7	1.58	(0.61, 4.13)	0.35	0.54
rs6060627 (BCL2L1)	CC	146	29	1 ^a				
	CT	173	22	0.61	(0.33, 1.12)	0.11		
	TT	52	5	0.49	(0.18, 1.35)	0.17	0.07	
Females only	rs2736100 (TERT)	CC	82	7	1 ^b			
		CA	127	21	1.93	(0.79, 4.76)	0.15	
		AA	72	9	1.47	(0.52, 4.16)	0.47	0.48
	rs1317082 (TERC)	AA	175	24	1 ^b			
		AG	104	13	0.91	(0.44, 1.87)	0.80	
		GG	17	5	2.12	(0.71, 6.27)	0.18	0.43
	rs2738783 (RTEL1)	TT	210	33	1 ^b			
		TG	85	9	0.67	(0.30, 1.45)	0.31	
		GG	7	2	1.68	(0.33, 8.59)	0.54	0.65
	rs10165485 (ACYP2)	TT	250	32	1 ^b			
		TC	44	11	1.93	(0.91, 4.12)	0.09	
		CC	6	0	N/A	N/A	N/A	N/A
	rs2320615 (NAF1)	GG	171	24	1 ^b			
		GA	110	17	1.10	(0.57, 2.15)	0.77	
		AA	22	3	1.0	(0.28, 3.61)	1.00	0.86
	rs9257445 (ZNF311)	GG	170	28	1 ^b			
		GC	115	10	0.53	(0.25, 1.13)	0.10	
		CC	16	6	2.27	(0.82, 6.29)	0.12	0.91
rs6060627 (BCL2L1)	CC	115	24	1 ^b				
	CT	144	17	0.56	(0.29, 1.10)	0.09		
	TT	44	3	0.33	(0.09, 1.14)	0.08	0.03	
Female only	rs2736100 (TERT)	CC	82	4	1 ^b			
		CA	127	11	1.77	(0.55, 5.76)	0.34	
		AA	72	6	1.71	(0.46, 6.32)	0.42	0.43
	rs1317082 (TERC)	AA	175	11	1 ^b			
		AG	104	11	1.68	(0.70, 4.01)	0.24	
		GG	17	4	3.74	(1.07, 13.05)	0.039	0.10
	rs2738783 (RTEL1)	TT	210	19	1 ^b			
		TG	85	8	1.04	(0.44, 2.48)	0.93	
		GG	7	0	N/A	N/A	N/A	N/A
	rs10165485 (ACYP2)	TT	250	22	1 ^b			
		TC	44	4	1.03	(0.34, 3.18)	0.96	
		CC	6	0	N/A	N/A	N/A	N/A
	rs2320615 (NAF1)	GG	171	15	1 ^b			
		GA	110	11	1.14	(0.51, 2.57)	0.75	
		AA	22	1	0.52	(0.07, 4.13)	0.54	0.83

Table 5. Continued

SNP	Genotype	Unaffected	Affected	OR	(95%CI)	P	P _{trend} ^c
rs9257445 (ZNF311)	GG	170	18	1 ^b			
	GC	115	6	0.49	(0.19, 1.28)	0.15	
rs6060627 (BCL2L1)	CC	16	3	1.77	(0.47, 6.66)	0.40	0.72
	CC	115	15	1 ^b			
	CT	144	10	0.53	(0.23, 1.23)	0.14	
	TT	44	2	0.35	(0.08, 1.59)	0.17	0.08

^aAdjusted with age at last interview/diagnosis and gender.

^bAdjusted with age at last interview/diagnosis.

^cP value of the trend test.

which could modulate this risk, as well as potential genotype-phenotype correlations.

Given the functional link between ATM and telomere maintenance, we hypothesized that TL and telomere-related SNPs could act as modifiers of cancer risk for HetAT subjects. Interestingly, we observed a sex-dependent effect of some SNPs on RTL in HetAT subjects. Indeed, the β coefficient for rs10165485, rs2320615 and rs9257445 was 2- to 5-fold higher in males than in females. Such a comparison in non-HetAT subjects was not possible due to the limited number of non-HetAT males in the study population (Supplementary Table 1). This difference in gender was not observed in the general population as reported in Codd et al. 2013 (29). We have no obvious explanation for this sex-dependent effect in ATM carriers, and larger studies are warranted to confirm this finding.

Regarding the possible modifier role of TL on cancer risk, we found that RTL, as measured with the Q-PCR assay, was not a risk factor for cancer susceptibility in our study population. In the SNP analysis, only rs6060627 in BCL2L1 showed association with overall cancer risk although this intronic SNP was not associated with TL. Therefore a mechanism other than telomere maintenance involving BCL2L1 could explain the observed association here. There are several limitations to this study that should be noted. First, the power to detect SNPs with small effect sizes may be low due to the relatively small number of cancer cases included in this study because of the uniqueness of the studied population. Indeed for SNP rs10165485 with a low MAF in controls (10%), the power of our study for detecting an association between this candidate SNP and cancer could reach 80% only for an OR of 2.5 or higher ($\alpha < 0.05$). For SNP rs2536100 with a MAF in controls of 50%, our study had a power of 90% for detecting an association with an OR equals 1.40 ($\alpha = 0.05$). Second, due to the limited number of cases, the main analyses were conducted considering all cancer sites together, which could mask association with some specific cancer sites. Indeed, risk of non-Hodgkin lymphoma (40), lung (18) and prostate (41) cancers have been associated with longer telomeres, whereas melanoma (22), thyroid (42) and ovarian (43) cancers have been associated with shorter telomeres. No significant association between BC risk and RTL was found here when sorting RTL into tertiles; moreover the association between ATM mutation status and cancer was not contingent with RTL (i.e. no statistical interaction was found). However, classifying subjects according to their ATM mutation status and Long vs. Medium/Short telomeres suggested that HetAT individuals with shorter telomeres have an increased overall cancer risk and an increased BC risk as compared to HetAT individuals with Long telomeres. This observation is in agreement with the hypothesis that shorter telomeres increased genomic instability, as observed in A-T children. Indeed, studies on lymphocytes and fibroblasts from A-T children have shown an accelerated

shortening of telomeres, a deregulation probably involved of the chromosome end-to-end fusion and genomic instability characteristic of the disorder (44,45). The accelerated telomere shortening ascertained for A-T children and HetAT cancer cases could be explained by the involvement of ATM in two mechanisms of telomere lengthening regulation. The first pathway involves the phosphorylation of TFR1 by ATM which is required to release TRF1 from telomeric DNA, ensuring the formation of replication forks that allow telomerase recruitment (12,46). In the second TRF1-independent pathway, ATM facilitates the telomerase assembly by phosphorylation of currently unknown proteins (9). Thus, in HetAT individuals a reduction in the quantity of functional ATM protein might cause a lower efficiency to induce replication fork formation and telomerase assembly, leading to an accelerated telomere shortening and an increased genomic instability. Despite these observations, no statistically significant interaction between ATM status and telomere length was found in cancer risk but observations are too few to exclude this possibility. The finding that unaffected ATM mutation carriers have longer TL than their non-carrier relatives was unexpected. In order to assess the possibility of a different dispersion of RTL measures in HetAT and non-HetAT subjects, we further explored median and mean RTL in the two groups according to age (Supplementary Figure 1). No difference in median RTL was seen between non-HetAT and HetAT subjects in each age group but the decline in mean RTL according to age was more marked for non-HetAT subjects than for HetAT subjects, which is in accordance with our observation that telomere shortening is slower in HetAT subjects (Figure 1).

One possible explanation for HetAT subjects having longer TL could be that the quantitative PCR assay used also amplifies extra-chromosomal telomeric DNA, thus explaining the surplus of telomere signals. Linear or circular extra-chromosomal telomeric DNA is characteristic of alternative lengthening of telomeres (ALT) a mechanism independent of the telomerase (47). The formation of such elements was observed in fibroblasts from A-T children (48). Hence, the ALT mechanism might participate in telomere lengthening in cells from HetAT subjects, as previously shown in BRCA2-/+ cell lines (49) and could open new lines of research into the function of ATM. This deregulation might be due to some environmental exposures like radiation or to other unidentified genetic factors involved in TL maintenance.

Taken together, our findings suggest that TL may be modulated by the carriage of an ATM mutation but it does not seem itself directly related to increased cancer risk in carriers. TL measurement alone is not a good marker for predicting cancer risk in A-T families or for predicting ATM mutation status of an individual.

Supplementary Material

Supplementary data are available at Carcinogenesis online.

Funding

This work was supported by Inserm and Ministère de la Recherche (01P0751 - 01P0752 - 01P0753 - 01P0754 - 01P0755), Electricité de France (conseil scientifique de Radioprotection d'EDF, grant EP 2002-03, EP 2004-03 RB 2016-22), Fondation de France (grants 2001009761 and 2005011201), La Ligue (grants PRE04/NA, PRE07/NA and PRE2015 LNCC/NA), La Ligue Comité du Maine et Loire, MGEN Union, ITMO Santé Publique d'AVIESAN (grant AAP12-COH-110), Institut Curie (CEST NC2013-015), Institut National du Cancer (grant INCa-9578) and the European Network of Excellence DoReMi (grant 249689). A.L.R. was the recipient of a fellowship from the Fondation ARC pour la recherche sur le cancer.

Acknowledgments

We are most grateful to the families who so willingly participated into the study. We acknowledge José-Arturo Londono Vallejo for helpful discussions regarding the set-up of experiments for measuring telomere length. We thank all the CoF-AT collaborating cancer clinics: Institut Curie, Paris: Dominique Stoppa-Lyonnet, Bruno Buecher, Antoine De Pauw and Sophie Lejeune-Dumoulin; Hôpital Arnaud de Villeneuve, Montpellier: Isabelle Coupier; ICM Val d'Aurel, Montpellier: Isabelle Coupier; CHU de Nîmes: Audrey Combès; Centre François Baclesse, Caen: Pascaline Berthet; Hôpital Saint-Joseph, Marseille: Hélène Zattara; Clinique Universitaire Saint-Luc, Bruxelles: Nicolas Janin and Karin Dahan; Centre Oscar Lambret, Lille: Philippe Vennin and Claude Adenis; Institut Bergonié, Bordeaux: Michel Longy; Centre Jean Perrin, Clermont-Ferrand: Jacques-Olivier Bay; Centre Jean-Paul Strauss, Strasbourg: Jean-Pierre Fricker; Centre Catherine de Sienne, Nantes: Alain Lortholary; CHU de Poitiers, Poitiers: Brigitte Gilbert-Dussardier; Hôpital d'Enfants, Dijon: Laurence Faivre and Caroline Jacquot; Centre Antoine Lacassagne, Nice: Marc Frenay; Institut Claudius Regaud, Toulouse: Laurence Gladiéff; CHU de Grenoble, Grenoble: Dominique Leroux; CHU de Lyon, Lyon: Gaétan Lesca; Gustave Roussy, Villejuif: Agnès Chompret[†]; Centre Léon Bérard, Lyon: Christine Lasset; Hotel-Dieu, Chambéry: James Lespinasse; CHI Toulon La Seyne-sur-Mer, Toulon: Xavier Tchiknavorian; Centre Eugène Marquis, Rennes: Catherine Dugast[†]; CHU du Mans, Le Mans: Dominique Martin-Coignard; CHRU Hôpital Caremeau, Nîmes: Jean Chiesa; Hôpital Universitaire Dupuytren, Limoges: Laurence Venat-Bouvet; Centre René Gauducheau, Nantes: Capucine Delnatte; Hôpital Porte-Madeleine, Orléans: Sonia Nizard; CHU de Nancy, Nancy: Bruno Leheup; Hôpital Sainte-Musse, Toulon: Patrick Collignon; Polyclinique Courlancy, Reims: Liliane Demange[†]; CHU de Besançon-Hôpital Jean Minjoz, Besançon: Pierre Rohrlrich; CHU Vaudois, Lausanne: Florence Fellmann; CHU d'Angers, Angers: Isabelle Pellier; Hôpitaux de Rouen, Rouen: Julie Tinat. We wish to thank more specifically our late colleague Josué Feingold who has been instrumental in the set-up of the CoF-AT study. [†]Deceased prematurely.

Conflict of Interest Statement: None declared.

Authors contribution

Conceived and designed the experiments: A.L.R., J.H., N.A., F.L. Performed the experiments: A.L.R., N.M., C.D.E. Analyzed the data: A.L.R., N.A., F.L. Contributed reagents/materials/analysis tools: J.L., C.D.E., A.L., M.G.D. Coordinated the CoF-AT and Retro-AT studies: D.S.L., N.A., J.H. Included CoF-AT participants, collected data: E.C., D.L.G., M.L. Invited A-T family members: G.L., D.L., L.G., C.A., M.F., L.F., B.G.-D., A.L., J.-P.F., K.D., J.-O.B., M.L., B.B.,

P.V.†, N.J., H.Z., P.B., A.C., I.C., D.S.L. Wrote the paper: A.L.R., F.L. All authors read and approved the final version of the manuscript.

References

- Swift, M. et al. (1976) Malignant neoplasms in the families of patients with ataxia-telangiectasia. *Cancer Res.*, 36, 209–215.
- Børresen, A.L. et al. (1990) Breast cancer and other cancers in Norwegian families with ataxia-telangiectasia. *Genes. Chromosomes Cancer*, 2, 339–340.
- Andrieu, N. et al. (2005) Ataxia-telangiectasia genes and breast cancer risk in a French family study. *J. Dairy Res.*, 72 Spec No, 73–80.
- Thompson, D. et al. (2005) Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl. Cancer Inst.*, 97, 813–822.
- Renwick, A. et al.; Breast Cancer Susceptibility Collaboration (UK). (2006) ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.*, 38, 873–875.
- Tavtigian, S.V. et al.; Australian Cancer Study; Breast Cancer Family Registries (BCFR); Kathleen Cunningham Foundation Consortium for Research into Familial Aspects of Breast Cancer (kConFab). (2009) Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am. J. Hum. Genet.*, 85, 427–446.
- Blackburn, E.H. (1991) Structure and function of telomeres. *Nature*, 350, 569–573.
- Cervantes, R.B. et al. (2002) Mechanisms of chromosome-end protection. *Curr. Opin. Cell Biol.*, 14, 351–356.
- Tong, A.S. et al. (2015) ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. *Cell Rep.*, 13, 1633–1646.
- Verdun, R.E. et al. (2006) The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell*, 127, 709–720.
- McKerlie, M. et al. (2012) ATM regulates proteasome-dependent subnuclear localization of TRF1, which is important for telomere maintenance. *Nucleic Acids Res.*, 40, 3975–3989.
- Wu, Y. et al. (2007) MRE11-RAD50-NBS1 and ATM function as co-mediators of TRF1 in telomere length control. *Nat. Struct. Mol. Biol.*, 14, 832–840.
- Sfeir, A. et al. (2012) Removal of shelterin reveals the telomere end-protection problem. *Science*, 336, 593–597.
- Harley, C.B. et al. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature*, 345, 458–460.
- Lindsey, J. et al. (1991) *In vivo* loss of telomeric repeats with age in humans. *Mutat. Res.*, 256, 45–48.
- DePinho, R.A. (2000) The age of cancer. *Nature*, 408, 248–254.
- Hou, L. et al. (2009) Telomere length in peripheral leukocyte DNA and gastric cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, 18, 3103–3109.
- Rode, L. et al. (2013) Short telomere length, lung function and chronic obstructive pulmonary disease in 46,396 individuals. *Thorax*, 68, 429–435.
- Broberg, K. et al. (2005) Constitutional short telomeres are strong genetic susceptibility markers for bladder cancer. *Carcinogenesis*, 26, 1263–1271.
- Risques, R.A. et al. (2007) Leukocyte telomere length predicts cancer risk in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.*, 16, 2649–2655.
- Lan, Q. et al. (2009) A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of non-Hodgkin lymphoma. *Clin. Cancer Res.*, 15, 7429–7433.
- Anic, G.M. et al. (2013) Telomere length and risk of melanoma, squamous cell carcinoma, and basal cell carcinoma. *Cancer Epidemiol.*, 37, 434–439.
- Shen, J. et al. (2009) Telomere length, oxidative damage, antioxidants and breast cancer risk. *Int. J. Cancer*, 124, 1637–1643.
- Qu, S. et al. (2013) Association of leukocyte telomere length with breast cancer risk: nested case-control findings from the Shanghai Women's Health Study. *Am. J. Epidemiol.*, 177, 617–624.
- Gramatges, M.M. et al. (2010) Longer relative telomere length in blood from women with sporadic and familial breast cancer compared with healthy controls. *Cancer Epidemiol. Biomarkers Prev.*, 19, 605–613.

26. Pellatt, A.J. et al. (2013) Telomere length, telomere-related genes, and breast cancer risk: the breast cancer health disparities study. *Genes. Chromosomes Cancer*, 52, 595–609.
27. Kim, S. et al. (2011) Telomere length in peripheral blood and breast cancer risk in a prospective case-cohort analysis: results from the Sister Study. *Cancer Causes Control*, 22, 1061–1066.
28. Pooley, K.A. et al.; EMBRACE. (2014) Lymphocyte telomere length is long in BRCA1 and BRCA2 mutation carriers regardless of cancer-affected status. *Cancer Epidemiol. Biomarkers Prev.*, 23, 1018–1024.
29. Codd, V. et al.; CARDIoGRAM consortium. (2013) Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.*, 45, 422–427, 427e1.
30. Pooley, K.A. et al. (2013) A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Hum. Mol. Genet.*, 22, 5056–5064.
31. Bojesen, S.E. et al.; Australian Cancer Study; Australian Ovarian Cancer Study; Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer (kConFab); Gene Environment Interaction and Breast Cancer (GENICA); Swedish Breast Cancer Study (SWE-BRCA); Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); Epidemiological study of BRCA1 & BRCA2 Mutation Carriers (EMBRACE); Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO). (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.*, 45, 371–384, 384e1.
32. Kojis, T.L. et al. (1991) The cytogenetics of ataxia telangiectasia. *Cancer Genet. Cytogenet.*, 56, 143–156.
33. Pandita, T.K. et al. (1995) Chromosome end associations, telomeres and telomerase activity in ataxia telangiectasia cells. *Cytogenet. Cell Genet.*, 71, 86–93.
34. Janin, N. et al. (1999) Breast cancer risk in ataxia telangiectasia (AT) heterozygotes: haplotype study in French AT families. *Br. J. Cancer*, 80, 1042–1045.
35. Cavaciuti, E. et al. (2005) Cancer risk according to type and location of ATM mutation in ataxia-telangiectasia families. *Genes. Chromosomes Cancer*, 42, 1–9.
36. Cawthon, R.M. (2002) Telomere measurement by quantitative PCR. *Nucleic Acids Res.*, 30, e47.
37. Gardner, M. et al.; Halcyon study team. (2014) Gender and telomere length: systematic review and meta-analysis. *Exp. Gerontol.*, 51, 15–27.
38. Lapham, K. et al. (2015) Automated Assay of Telomere Length Measurement and Informatics for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*, 200, 1061–1072.
39. Burke, L.S. et al. (2013) Telomere length and the risk of cutaneous malignant melanoma in melanoma-prone families with and without CDKN2A mutations. *PLoS One*, 8, e71121.
40. Machiela, M.J. et al. (2016) Genetically predicted longer telomere length is associated with increased risk of B-cell lymphoma subtypes. *Hum. Mol. Genet.*, 25, 1663–1676.
41. Julin, B. et al. (2015) Circulating leukocyte telomere length and risk of overall and aggressive prostate cancer. *Br. J. Cancer*, 112, 769–776.
42. He, M. et al. (2013) Telomere length is shorter in affected members of families with familial nonmedullary thyroid cancer. *Thyroid*, 23, 301–307.
43. Mirabello, L. et al. (2010) Leukocyte telomere length in a population-based case-control study of ovarian cancer: a pilot study. *Cancer Causes Control*, 21, 77–82.
44. Metcalfe, J.A. et al. (1996) Accelerated telomere shortening in ataxia telangiectasia. *Nat. Genet.*, 13, 350–353.
45. Xia, S.J. et al. (1996) Reduced telomere length in ataxia-telangiectasia fibroblasts. *Mutat. Res.*, 364, 1–11.
46. Sfeir, A. et al. (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*, 138, 90–103.
47. Henson, J.D. et al. (2002) Alternative lengthening of telomeres in mammalian cells. *Oncogene*, 21, 598–610.
48. Hande, M.P. et al. (2001) Extra-chromosomal telomeric DNA in cells from Atm(-/-) mice and patients with ataxia-telangiectasia. *Hum. Mol. Genet.*, 10, 519–528.
49. Bodvardsdottir, S.K. et al. (2012) Dysfunctional telomeres in human BRCA2 mutated breast tumors and cell lines. *Mutat. Res.*, 729, 90–99.

7. Etude complémentaire : contribution de SNPs associés au risque de cancer et en particulier au risque de cancer du sein dans les familles A-T.

En plus de l'étude de la contribution des SNPs associés à la longueur des télomères dans le risque de cancer dans les familles A-T présentée dans l'Article 1, j'ai examiné le rôle de 15 SNPs associés au risque de cancer du sein dans la population générale dans le même échantillon de familles A-T en suivant la même stratégie d'analyse. En particulier, étant donné la petite taille de notre échantillon (pour rappel 167 femmes HetAT et 157 non-AT) (Tableau 14), j'ai dans un premier temps examiné la contribution de ces 15 SNPs dans le risque de cancer en prenant en compte l'ensemble des cancers diagnostiqués dans les familles A-T étudiés chez les sujets dont l'ADN était disponibles. Les résultats des tests d'association montrent une augmentation significative du risque de cancer pour le SNP rs2380295 (*ANKRD16*) (OR=5.1 ; 95%IC (1.36 ; 19.12) ; p -trend=0.02) (Tableau 15).

J'ai ensuite stratifié les patientes en fonction de la localisation du cancer. Le cancer du sein étant le type de cancer plus fréquent dans notre population d'étude j'ai analysé l'association entre les 15 SNPs et ce type de cancer (Tableau 15). Les résultats des tests d'association montrent qu'aucun SNP n'est associé avec le risque de cancer du sein dans notre population

Le SNP rs2380295 (*ANKRD16*) présente une association avec le risque de cancer lorsque que toutes les femmes atteintes d'un cancer sont prises en compte dans l'analyse, mais cette association n'est pas retrouvée quand l'analyse est stratifiée sur les cas de cancer du sein. Le calcul du test de puissance suggère que nous aurions pu détecter une association de 1,8 avec une puissance de 98% (Tableau 16). Ce SNP n'est donc pas associé spécifiquement avec le cancer du sein chez les femmes des familles A-T.

Tableau 14 : Distribution des cas de cancer chez les femmes participantes en fonction du statut mutationnel pour ATM

	Femmes indemnes (N=384)	Femmes atteintes d'un cancer (N=46)	Femmes atteintes d'un cancer du sein (N=29)
Non-HetAT	157	13	7
HetAT	167	33	22

Tableau 15 : Association entre les SNPs associés au cancer du sein et le risque de cancer chez les participantes de CoF-AT et Retro-AT.

SNP	Génotype	Femmes indemnes	Femmes atteintes d'un cancer (tous cancers)	OR*	(95%CI)	P	P _{trend} ^b
rs2046210 (<i>ESR1</i>)	GG	107	12	1 ^a			
	GA	152	25	1.44	(0.69, 3.01)	0.33	
	AA	51	7	1.22	(0.45, 3.29)	0.70	0.84
rs2380295 (<i>ANKRD16</i>)	CC	83	3	1 ^a			
	CT	165	29	4.87	(1.44, 16.44)	0.01	
	TT	61	11	5.1	(1.36, 19.12)	0.02	0.02
rs4973768 (<i>SLC4A7</i>)	CC	75	18	1 ^a			
	CT	148	13	0.36	(0.17, 0.78)	0.01	
	TT	87	13	0.61	(0.28, 1.33)	0.21	0.18
rs10995190 (<i>ZNF365</i>)	GG	209	31	1 ^a			
	GA	94	13	0.92	(0.46, 1.84)	0.83	
	AA	6	0	N/A	N/A	N/A	0.56
rs13387042 (<i>IGFPB5</i>)	AA	94	15	1 ^a			
	AG	149	17	0.71	(0.34, 1.49)	0.38	
	GG	62	12	1.22	(0.52, 2.73)	0.67	0.68
rs13281615 (<i>CASC8</i>)	AA	83	20	1 ^a			
	AG	161	18	0.47	(0.23, 0.93)	0.03	
	GG	53	6	0.48	(0.18, 1.29)	0.15	0.06
rs3803662 (<i>TNRC9</i>)	GG	154	20	1 ^a			
	GA	126	19	1.2	(0.61, 2.36)	0.60	
	AA	24	5	1.56	(0.53, 4.57)	0.42	0.39
rs2981578 (<i>FGFR2</i>)	GG	71	11	1 ^b			
	GC	146	21	0.92	(0.42, 2.02)	0.85	
	CC	89	12	0.87	(0.36, 2.08)	0.75	0.75
rs1045485 (<i>CAPS8</i>)	GG	238	35	1 ^a			
	GC	60	8	0.92	(0.40, 3.28)	0.83	
	CC	5	1	1.46	(0.16, 13.05)	0.73	0.99
rs889312 (<i>MAP3K1</i>)	AA	132	22	1 ^b			
	AC	139	17	0.74	(0.38, 1.46)	0.39	
	CC	30	5	0.99	(0.35, 2.84)	0.99	0.66
rs10941679 (<i>FGF10</i>)	AA	161	20	1 ^b			
	AG	114	19	1.33	(0.68, 2.67)	0.40	
	GG	26	5	1.52	(0.52, 4.41)	0.44	0.33
rs1011970 (<i>CDKN2A</i>)	GG	205	28	1 ^b			
	GT	86	15	1.34	(0.67, 2.67)	0.40	
	TT	12	1	0.66	(0.08, 5.33)	0.70	0.69
rs704010 (<i>ZMIZ1</i>)	CC	127	19	1 ^b			
	CT	142	16	0.76	(0.38, 1.55)	0.45	
	TT	41	9	1.48	(0.62, 3.52)	0.38	0.64

rs11249433 (<i>EMBP1</i>)	AA	101	17	1 ^b					
	AG	153	19	0.73	(0.36, 1.48)	0.39			
	GG	49	8	0.98	(0.39, 2.43)	0.96	0.78		
rs3817198 (<i>LSP1</i>)	TT	134	17	1 ^b					
	TC	147	22	1.17	(0.6, 2.3)	0.65			
	CC	25	5	1.55	(0.52, 4.59)	0.43	0.43		
(cancer du sein)									
rs2046210 (<i>ESR1</i>)	GG	107	7	1 ^a					
	GA	152	14	1.41	(0.55, 3.62)	0.48			
	AA	51	6	1.8	(0.58, 5.63)	0.31	0.30		
rs2380295 (<i>ANKRD16</i>)	CC	83	2	1 ^a					
	CT	165	19	4.78	(1.09, 21.01)	0.04			
	TT	61	6	4.1	(0.8, 21.19)	0.09	0.10		
rs4973768 (<i>SLC4A7</i>)	CC	75	12	1 ^a					
	CT	148	9	0.38	(0.15, 0.94)	0.04			
	TT	87	6	0.43	(0.15, 2.2)	0.11	0.08		
rs10995190 (<i>ZNF365</i>)	GG	209	18	1 ^a					
	GA	94	8	0.94	(0.40, 2.22)	0.88			
	AA	6	0	N/A	N/A	N/A	0.65		
rs13387042 (<i>IGFBP5</i>)	AA	94	9	1 ^a					
	AG	149	11	0.77	(0.31, 1.93)	0.38			
	GG	62	7	1.18	(0.42, 3.33)	0.76	0.83		
rs13281615 (<i>CASC8</i>)	AA	83	10	1 ^a					
	AG	161	12	0.62	(0.26, 1.49)	0.29			
	GG	53	5	0.78	(0.25, 2.44)	0.67	0.54		
rs3803662 (<i>TNRC9</i>)	GG	154	10	1 ^a					
	GA	126	12	1.46	(0.61, 3.51)	0.60			
	AA	24	5	3.23	(1.01, 10.32)	0.05	0.07		
rs2981578 (<i>FGFR2</i>)	GG	71	7	1 ^a					
	GC	146	14	0.97	(0.38, 2.52)	0.83			
	CC	89	6	0.68	(0.22, 2.12)	0.75	0.51		
rs1045485 (<i>CAPS8</i>)	GG	238	20	1 ^a					
	GC	60	6	1.19	(0.46, 3.1)	0.83			
	CC	5	1	2.4	(0.26, 21.98)	0.44	0.49		
rs889312 (<i>MAP3K1</i>)	AA	132	13	1 ^a					
	AC	139	12	0.88	(0.39, 1.99)	0.76			
	CC	30	2	0.68	(0.15, 3.16)	0.62	0.60		
rs10941679 (<i>FGF10</i>)	AA	161	10	1 ^a					
	AG	114	13	1.84	(0.78, 4.34)	0.17			
	GG	26	4	2.49	(0.72, 8.55)	0.15	0.09		
rs1011970 (<i>CDKN2A</i>)	GG	205	18	1 ^a					
	GT	86	9	1.19	(0.51, 2.79)	0.68			
	TT	12	0	N/A	N/A	N/A	0.80		
rs704010 (<i>ZMIZ1</i>)	CC	127	12	1 ^a					
	CT	142	9	0.67	(0.27, 1.65)	0.38			
	TT	41	6	1.55	(0.55, 4.39)	0.41	0.69		

rs11249433 (<i>EMBP1</i>)	AA	101	11	1 ^a				
	AG	153	11	0.66	(0.28, 1.58)	0.39		
	GG	49	5	0.94	(0.31, 2.85)	0.91	0.71	
rs3817198 (<i>LSP1</i>)	TT	134	8	1 ^a				
	TC	147	14	1.6	(0.65, 3.93)	0.65		
	CC	25	5	3.4	(1.01, 11.14)	0.05	0.06	

^a Ajustement sur l'âge au diagnostic pour les cas de cancer et sur l'âge aux dernières nouvelles pour les sujets indemnes.

^b p-value pour le test de tendance.

Tableau 16 : Puissance pour l'étude d'association entre le risque de cancer et les SNPs étudiés.

OR attendu		1,05	1,2	1,5	1,8
Fréquence de l'allèle à risque dans la population témoin	0,05	0,06	0,05	0,08	0,12
	0,10	0,05	0,04	0,11	0,23
	0,15	0,04	0,05	0,17	0,33
	0,20	0,03	0,06	0,21	0,45
	0,25	0,03	0,07	0,28	0,56
	0,30	0,03	0,08	0,33	0,68
	0,35	0,03	0,09	0,42	0,80
	0,40	0,03	0,10	0,48	0,91
	0,45	0,03	0,11	0,60	0,98
	0,50	0,03	0,13	0,69	0,99

OR : Odds-Ratio. Les résultats indiquent la probabilité de détecter une association entre SNP et cancer du sein en fonction de la fréquence de l'allèle à risque dans la population témoin et de l'odds ratio attendu.

8. DISCUSSION

Les travaux réalisés dans cette première partie de la thèse ont permis de mieux caractériser la population d'étude CoF-AT. En effet, j'ai confirmé la présence d'une mutation pathogène dans d'autres gènes associés à un risque élevé de cancer du sein (deux familles ségrégent une mutation *BRCA1* et deux autres familles ségrégent une mutation *BRCA2*) et ai mis en évidence des variants délétères dans des gènes de prédisposition à un autre type de cancer (variant tronquant de *MSH6* et de *MUTYH*). Les quatre femmes porteuses d'une mutations de *BRCA1* ou *BRCA2*, ainsi que la participante porteuse du variant de *MSH6*, ont été écartées des analyses suivantes visant à rechercher des facteurs génétiques constitutionnels modificateurs du risque de cancer dans les familles A-T. De plus, ces résultats m'ont permis de vérifier que les tumeurs du sein collectées pour réaliser le deuxième volet de la thèse n'étaient pas liées à la présence d'un autre variant délétère dans un autre gène de prédisposition au cancer du sein afin de ne pas inclure de biais dans l'analyse.

L'analyse de la longueur des télomères dans les familles A-T révèle que les individus HetAT ont des télomères plus long que leurs apparentés non-HetAT. La protéine ATM permettant l'accès à la télomérase via la phosphorylation de TPP1, la présence d'une copie défectueuse d'*ATM* laissait présager que les individus HetAT auraient des télomères plus courts que les individus non-HetAT. Ces résultats suggèrent qu'un autre mécanisme encore inconnu de la maintenance de la longueur des télomères est utilisé par la cellule chez les individus HetAT. Bien que la longueur des télomères ne semble pas associée au risque de cancer dans nos familles, lorsque l'on stratifie les individus selon la présence de la mutation familiale d'*ATM* et selon la longueur des télomères, on remarque que le risque de cancer augmente chez les individus HetAT ayant des télomères courts.

L'un des avantages de notre étude est la sélection des familles *ATM* participant à CoF-AT et Retro-AT, évitant ainsi les problèmes de classification des variants, puisqu'ils peuvent être donc tous considérés comme pathogène pour l'A-T. L'autre avantage est la réalisation du prélèvement sanguin avant le diagnostic de cancer, ce qui évite les biais dus aux traitements anti-cancéreux ayant un impact sur la longueur des télomères. Néanmoins, la limite de ce travail est la petite taille de l'échantillon. L'analyse des facteurs modificateurs du risque n'a pas pu être réalisée sur tous les types de cancer, ni sur les hommes spécifiquement, ces derniers n'étant pas éligibles à l'inclusion dans l'étude CoF-AT.

Le génotypage des sept SNPs associés à longueur des télomères dans la population générale montre que seul le SNP rs9257445 (*ZNF311*) est associé à la longueur des télomères dans notre population d'étude. Cela confirme que ce SNP peut être utilisé comme biomarqueur de la longueur des télomères dans les études d'épidémiologie génétique.

**Partie III : Caractérisation histologique et
moléculaire des tumeurs du sein
développées par les porteurs d'une ou
deux copies mutées d'*ATM*.**

1. Contexte de l'étude

Comme nous l'avons vu dans l'introduction, la prise en charge des patients repose désormais sur une description complète de la tumeur, à la fois morphologique et moléculaire.

De plus en plus d'études tendent à préciser les caractéristiques des tumeurs du sein survenant dans le cadre de prédispositions génétiques.

1.1 Caractéristiques morphologiques et moléculaires des tumeurs du sein développées par les porteurs *BRCA1*

1.1.1 Caractéristiques morphologiques

Les tumeurs du sein développées par les porteurs d'une mutation constitutionnelle de *BRCA1* sont à ce jour les tumeurs les plus étudiées et les mieux caractérisées. D'un point de vue morphologique, les tumeurs développées par les porteurs d'une mutation *BRCA1* (« tumeurs *BRCA1* ») sont souvent des carcinomes invasifs médullaires²⁹⁹⁻³⁰¹ caractérisés par une délimitation nette de la tumeur, une architecture syncytiale (cellules agencées en travées), un stroma lymphocytaire, des atypies nucléaires modérées à marquées, et une absence de différenciation glandulaire³⁰². Les tumeurs *BRCA1* sont de haut grade (grade III)³⁰³⁻³⁰⁵ avec un index mitotique élevé traduisant leur capacité de prolifération³⁰³⁻³⁰⁶.

Quatre-vingt pour cent des tumeurs *BRCA1* sont des tumeurs dites « triple-négatives » qui n'expriment pas les récepteurs aux œstrogènes, à la progestérone, ni la protéine HER2³⁰⁶⁻³⁰⁸. Les tumeurs *BRCA1* exprimant les récepteurs aux œstrogènes sont souvent diagnostiquées chez les patients plus âgés^{108, 305}. Par ailleurs, environ 10% des tumeurs *BRCA1* expriment la protéine HER2³⁰⁹.

1.1.2 Caractéristiques moléculaires

Selon l'hypothèse des « deux-coups » de Knudson, le cancer serait le résultat de l'accumulation de mutations au niveau de l'ADN³¹⁰. En 1994, Neuhausen et Marshall proposent que le premier « coup » serait la mutation constitutionnelle et que le deuxième « coup » serait la perte de l'allèle normal. A partir de l'étude 14 tumeurs du sein et de l'ovaire *BRCA1*, ils confirment leur hypothèse en montrant que l'allèle normal de *BRCA1* est perdu dans 13 des 14 tumeurs étudiées³¹¹. Depuis, plusieurs autres études ont confirmé la prévalence de la perte de l'allèle normal dans les tumeurs *BRCA1*, confirmant ainsi le rôle de gène suppresseur de tumeur de *BRCA1*³¹²⁻³¹⁴.

Plus tard, l'analyse du transcriptome des tumeurs *BRCA1* à partir de puces à ADN a montré que ces tumeurs expriment les cytokératines CK5/6, CK14, CK17 et EGFR caractéristiques du sous-type moléculaire basal-like^{39, 315, 316}. Ce résultat est en adéquation avec la classification déterminée par

Perou et coll., les tumeurs basal-like faisant partie du groupe des tumeurs n'exprimant pas les récepteurs aux œstrogènes, ni les gènes régulés par cette hormones tel que la cycline D1 ou bcl2^{26, 317}. Les tumeurs *BRCA1* exprimant les récepteurs aux œstrogènes sont quant à elles de sous-type luminal.

Du point de vue des variations du nombre de copies, Tirkkonen et coll. montrent dès 1997 par hybridation fluorescente *in situ* (FISH) que les tumeurs mammaires *BRCA1* sont deux fois plus remaniées que les tumeurs des cas sporadiques. Les pertes des loci 2q, 4p, 4q, 5q et 12q apparaissent comme spécifiques aux tumeurs *BRCA1*³¹⁸. Par la suite, grâce aux données de génomique comparative *in situ* (CGH) et à la mise en évidence de sous-types intrinsèques, les pertes des locis 4p, 4q et 5q ont été associées au sous-type moléculaire basal-like.

La mise en évidence d'un sous-type intrinsèque associé aux tumeurs *BRCA1* ne permettait cependant pas de prédire parmi les tumeurs basales-like lesquelles étaient associées à une perte de fonction de *BRCA1*³¹⁹. En 2012, Popova et coll. cherchent à identifier des biomarqueurs permettant de prédire l'inactivation biallélique de *BRCA1* au sein du sous-groupe des tumeurs basal-like³²⁰. Pour ce faire, les auteurs évaluent les variations du nombre de copies du génome tumoral de porteuses de mutation dans *BRCA1* grâce à une puce à ADN ciblant des SNPs fréquents sur tout le génome (puce CytoScan d'Affymetrix) et analysent les données avec la méthode GAP (Genomic Alteration Print) qu'ils ont développée³²¹. Cet outil permet de déterminer à partir des données de puces à ADN la ploïdie de la tumeur, le nombre de copies des segments génomiques, l'hétérogénéité tumorale ainsi que la contamination des échantillons par l'ADN normal. La comparaison des profils génomiques obtenus par cette méthode à partir de 23 tumeurs *BRCA1* et 42 tumeurs de patientes non-porteuses d'une mutation constitutionnelle dans les gènes *BRCA1* ou *BRCA2* a permis de montrer que l'inactivation biallélique de *BRCA1* peut être prédite à partir du nombre de cassures chromosomiques de plus de 10Mb entre deux régions adjacentes (« LST » pour Large-scale State Transition). En effet, les auteurs montrent que les tumeurs n'exprimant pas *BRCA1* présentent plus de 15 LST dans les tumeurs pseudo-diploïdes, ou au moins 20 LST dans les tumeurs pseudo-tétraploïdes. L'étude révèle également qu'environ 80% des tumeurs *BRCA1* sont pseudo-diploïdes alors que les tumeurs des non-porteurs de la série témoin sont pseudo-tétraploïdes. La série de réplication composée de 55 tumeurs basales-like triple-négatives a confirmé ces résultats. Cette signature est nommée « BRCAness » ou « HR déficiency » (Déficiency de la réparation des cassures de l'ADN par recombinaison homologue).

Au niveau des altérations somatiques de type SNV, ou petite insertion ou délétion (indel), la prévalence des mutations dans les gènes *TP53* et *PTEN* a été rapportée dans plusieurs études^{31, 322}.

³²³. Parmi les 560 tumeurs du sein séquencées par l'équipe de MR. Stratton au Sanger Institute, 35 ont une mutation constitutionnelle ou somatique délétère de *BRCA1* et 16 autres tumeurs présentent une inactivation du gène du à la méthylation du promoteur de *BRCA1*. Les résultats suggèrent que les tumeurs *BRCA1* n'ont pas un profil mutationnel homogène, certaines étant caractérisées par un fort taux de substitutions accompagnées de délétions de plus de 3pb, alors que d'autres sont caractérisées par des doubles substitutions CC>AA ou de grand réarrangements³⁴. Les résultats ne disent pas s'il existe une différence de signature entre les tumeurs venant des porteurs *BRCA1* et celles venant de non-porteurs avec une mutation somatique délétère de ce gène.

1.2 Caractéristiques histologiques et moléculaires des tumeurs du sein des porteurs *BRCA2*

Le gène *BRCA2* étant impliqué, comme *BRCA1*, dans la réparation des dommages à l'ADN, l'hypothèse que les tumeurs développées par les porteurs *BRCA2* présentaient les mêmes caractéristiques histologiques et moléculaires que les tumeurs *BRCA1* a émergé.

1.2.1 Caractéristiques morphologiques

Les tumeurs *BRCA2* sont majoritairement des carcinomes canaux infiltrants NST^{309, 315} de grade intermédiaire à haut (grade II-III) avec un index mitotique élevé et peu de formation de tubules³¹⁵. Ces tumeurs expriment les récepteurs aux œstrogènes^{299, 309, 315, 324} mais n'expriment généralement pas la protéine HER2^{309, 315}. Les tumeurs *BRCA2* ne présentent donc pas de similitudes morphologiques avec les tumeurs *BRCA1*. Par ailleurs, l'atteinte des marges d'exérèse est plus fréquente dans les tumeurs *BRCA2*³¹⁵ que dans les tumeurs de cas sporadiques.

1.2.2 Caractéristiques moléculaires

Dès 1999, Collins et *coll.* montraient à partir de 9 tumeurs du sein *BRCA2*, que l'expression du gène était totalement inactivée dans 8 tumeurs à cause d'une perte de l'allèle normal³²⁵. En 2016, Nick-Zainal et *al.* confirmaient ce résultat dans une plus large série de tumeurs *BRCA2* (N=39)³⁴.

L'équipe de Tirkkonen et *coll.* qui avait par ailleurs mis en évidence par FISH des altérations spécifiques aux tumeurs *BRCA1*, a également rapporté dans les tumeurs *BRCA2* une prévalence des pertes des loci 13q et 6q accompagnées de gain aux loci 17q22-q24 et 20q13³¹⁸. Quelques années plus tard, van Beers et *al.* comparaient par CGH les profils génomiques de 36 tumeurs *BRCA1*, 25 tumeurs *BRCA2* et 30 tumeurs de cas sporadiques³²⁶. Le profil génomique des tumeurs *BRCA2* se révèle similaire à celui des cas sporadiques et non à celui des tumeurs *BRCA1*. Cette différence au niveau des profils génomiques se reflète au niveau du sous-type moléculaire intrinsèque puisque les

tumeurs *BRCA2* sont majoritairement lumineales A, et non basales-like comme les tumeurs *BRCA1*^{315, 327}. Par ailleurs, une perte de longs fragments chromosomiques a été mise en évidence à partir de l'analyse CGH de 18 tumeurs *BRCA2*³²⁷. Ce résultat est supporté par l'étude de Pecuchet et coll. trois en plus tard, comparant à partir d'une puce à ADN et de la méthode GAP, les profils CNV de 30 tumeurs RO+ de porteurs hétérozygotes *BRCA2* et de 215 tumeurs RO+ de cas sporadiques³²⁸. Cette étude montre que les tumeurs *BRCA2* ont un nombre plus important de cassures chromosomiques de plus de 3Mb que les tumeurs des cas sporadiques. Les tumeurs *BRCA2* sont donc, comme les tumeurs *BRCA1*, associée à la signature « *Homologous Recombiantion deficiency* » (HRD : Recombinaison homologue déficiente) De plus, la cooccurrence des pertes aux loci 13q13 et 14q32 est prédictive, dans les tumeurs RO+, de la présence d'une mutation constitutionnelle de *BRCA2* avec 90% de sensibilité³²⁸. La perte des loci 13q et 14q a également été rapportée dans une étude portant sur l'analyse de l'expression des gènes dans 27 tumeurs³²⁹.

Des résultats contradictoires ont été rapportés quant à ploidie des tumeurs *BRCA2*. En effet, une étude a rapporté que les tumeurs *BRCA2* sont plus souvent tétraploïdes que les tumeurs des cas sporadiques³³⁰, alors que d'autres études rapportent une diploïdie des tumeurs *BRCA2*^{320, 328}.

L'ensemble de ces résultats montre que malgré la présence de la signature HRD dans les tumeurs *BRCA2*, la morphologie et les caractéristiques moléculaires de ces tumeurs diffèrent des tumeurs *BRCA1*.

1.3 Caractéristiques histologiques et moléculaires des tumeurs du sein de porteurs d'une mutation dans d'autres gènes de prédisposition au cancer du sein.

La mise en évidence d'une signature HRD associée à la perte de fonction de *BRCA1* ou *BRCA2* a soulevé la question de la présence de cette signature dans les tumeurs développées par les porteurs d'une mutation dans d'autres gènes impliqués dans la réparation des dommages à l'ADN et prédisposant au cancer du sein. De plus, la mise en évidence de caractéristiques spécifiques aux tumeurs associées à d'autres gènes de prédisposition au cancer du sein pourrait servir de biomarqueurs pour identifier les porteurs d'une mutation constitutionnelle et personnaliser le suivi et le traitement de ces patients.

1.3.1 Tumeurs *PALB2*

Comme nous l'avons vu dans l'introduction de ce mémoire, la recherche de mutations constitutionnelles dans le gène *PALB2* a récemment été introduite dans les tests de diagnostic

moléculaire pour les familles hautement prédisposées au cancer du sein et de l'ovaire en parallèle de *BRCA1* et *BRCA2*. En raison souvent du petit échantillonnage des études, la morphologie des tumeurs associées à *PALB2* a fait l'objet de controverses pendant plusieurs années. Certaines études rapportent une prévalence de tumeurs RO+ et d'autres une prévalence des tumeurs triple-négatives^{109, 331, 332}. En 2015, Cybulski et coll. analysent la morphologie de 116 tumeurs de porteurs de mutations tronquantes de *PALB2* (mutations 509_510delGA ou 172_175delTTGT) et montrent que ces tumeurs sont plus triples-négatives (34%) que les tumeurs de non porteurs de mutation *PALB2* (14%)³³³.

Par ailleurs, Teo et coll. ont comparé la morphologie de 28 tumeurs *PALB2* à celle de 770 tumeurs de femmes non porteuses de mutations dans *BRCA1* et *BRCA2* et ont montré que la morphologie des tumeurs *PALB2* ne ressemblent pas à celle des tumeurs *BRCA1* et *BRCA2* mais qu'elles se distinguent par un plus fort taux de sclérose³³². La majorité des tumeurs analysées dans cette étude provenant de femmes porteuses de la mutation c.3113G>A (24/28 tumeurs), les auteurs n'ont pas pu conclure si la sclérose était liée en particulier à cette mutation ou aux tumeurs *PALB2*.

A ce jour seule une étude canadienne a recherché le LOH de l'allèle normal dans les tumeurs *PALB2* dans 5 tumeurs de porteuses de mutation tronquantes ou faux-sens de *PALB2*. Le LOH n'a été trouvé que dans 2 tumeurs provenant d'une même patiente porteuse de la mutation c.3507_3508delTC (p.H1170Ffs*19), la petite taille de l'échantillon ne permet cependant pas de conclure quant à la fréquence du LOH au locus 16p12.1 contenant *PALB2* dans ces tumeurs³³⁴.

1.3.2 Tumeurs *CHEK2*

Une étude polonaise portant sur les tumeurs de porteurs d'une mutation tronquante ou faux-sens de *CHEK2* montre qu'il y aurait une corrélation génotype-phénotype³³⁵. En effet, les tumeurs développées par les femmes porteuses d'une mutation tronquante de *CHEK2* (c.1100delC, c.444+1G>A et c.5395) sont de sous-type luminal B (18 tumeurs sur 28 testées) alors que les tumeurs développées par les femmes porteuses de la mutation faux-sens c.470T>C (p.Ile157Thr) sont de sous-type luminal A (72 tumeurs sur 88 testées) et donc de meilleur pronostic³³⁵. De plus, la mutation c.470T>C a été associée au carcinome lobulaire³³⁶.

Par ailleurs, une autre étude portant sur 18 tumeurs *CHEK2* montre que seulement 17% des tumeurs présentent une perte de l'allèle normal, suggérant que *CHEK2* intervient dans la prédisposition au cancer du sein principalement par un mécanisme d'haplo-insuffisance³³⁷. Ces observations ont été rapportées dans une série de 26 tumeurs développées par des porteurs de la mutation fondatrice 1100delC³³⁸. L'analyse des CNV de ces 26 tumeurs révèle une hétérogénéité des profils génomiques : un groupe est caractérisé par des gains des loci 1q, 8q et 16p et des pertes des loci 8p et 16q tandis

que le deuxième groupe présente des amplifications focales de *ERBB2* et est caractérisé par une forte infiltration des lymphocytes.

D'un point de vue prédictif, bien que les tumeurs *CHEK2* soient majoritairement RO+, il ne semble pas que l'utilisation des statuts hormonaux comme bio-marqueurs soit discriminant par rapport aux tumeurs de non-porteurs de mutation de ce gène³³⁹.

1.3.3 Tumeurs *RAD51C*

Les tumeurs développées par les porteurs d'une mutation constitutionnelle de *RAD51C* ont des caractéristiques histologiques associées à un pronostic favorable¹⁴⁶. En effet, une étude réalisée sur 10 tumeurs du sein montre qu'elles sont majoritairement de type canalaire sans atypies particulières et présentent un phénotype RO+, HER2-. L'étude de Gevensleben et coll. confirme ces observations en définissant le sous-type moléculaire de 22 tumeurs du sein comme luminal A³⁴⁰. De plus, cette étude montre que les tumeurs de patients porteurs d'un variant délétère de *RAD51C* sont de grade intermédiaire (grade II) alors que les tumeurs de porteurs de variants faux-sens de signification clinique inconnue sont de grade plus élevé (grade III). Par ailleurs, on distingue les tumeurs *RAD51C* des tumeurs *BRCA2* grâce au Ki-67, significativement plus élevé dans les tumeurs *BRCA2*. Une autre observation intéressante est que plusieurs études ont rapporté des tumeurs triple-négatives^{146, 331, 340, 341}, ce qui pourrait être lié à la perte de fonction du gène³³¹.

1.3.4 Tumeurs *NBN*

Peu d'études ont décrit les caractéristiques histologiques et moléculaires des tumeurs du sein de porteurs d'un variant de *NBN* (*NBS1*). Une étude polonaise sur des patients porteurs hétérozygotes de la mutation c.657_661delACAAA (p.Lys219Asnfs) de ce gène décrit cinq tumeurs du sein comme étant canalaire de grade intermédiaire à élevé (II-III) et toutes présentant une perte de l'allèle normale³⁴². Le LOH de l'allèle normal a également été trouvé chez des 7 des 8 cas de cancer de la prostate porteurs de la même mutation³⁴³. Cependant, ces résultats ne sont pas retrouvés dans une autre étude portant sur 4 tumeurs du sein, une seule de ces tumeurs montrant une perte de l'allèle normal du gène³³⁷.

Comparées aux tumeurs de patients non-porteurs de mutation dans ce gène, aucune différence n'a été mise en évidence entre le statut des récepteurs hormonaux, le grade histologique et de l'état d'invasion ganglionnaire³⁴⁴.

1.3.5 Tumeurs *CDH1* et carcinomes lobulaires

Les tumeurs du sein de porteurs d'une mutation délétère de *CDH1* sont majoritairement de sous-

type lobulaire³⁴⁵. Une étude a montré qu'il n'y avait pas de différence au niveau clinique et histologique entre les carcinomes lobulaires *CDH1*-mutés et les carcinomes lobulaires *CDH1* non mutés³⁴⁶. De part leur grade histologique (grade I) et un index mitotique faible ainsi qu'une leur faible prolifération cellulaire, les tumeurs lobulaires sont associées à un bon pronostic. Il a cependant été observé un plus fort taux de tumeurs multifocales³⁴⁷ ou de cas bilatéraux³⁴⁸.

Au niveau moléculaire, les carcinomes lobulaires sont associés au sous-type intrinsèque luminal A. Deux études récentes ont mis en évidence dans 630³⁴⁹ et 127³⁵⁰ carcinomes lobulaires respectivement la présence de mutations somatiques de *CDH1*, *FOXA1*, *GATA3* ainsi que trois gènes impliqués dans la voie du phosphatidylinositol 3-kinase, *PIK3CA*, *PTEN* et *AKT1*. En comparant les données de séquençage du génome des carcinomes lobulaires aux données publiques du *Cancer Genome Atlas* (TCGA), il a été montré qu'en plus des mutations de *CDH1*, les mutations des gènes *PTEN*, *FOXA1* et *TBX3* ainsi que l'amplification focales du locus 6q25 (*ESR1*) sont des bio-marqueurs de ce sous-type histologique.

D'un point de vue clinique, les amplifications de *ESR1* ont été plus fréquemment mises en évidence chez des patients âgés, et les mutations de *PIK3CA* sont considérées comme un bio-marqueur d'une faible prolifération tumorale. De plus, les mutations somatiques des gènes *PTEN* et *TBX3* sont associées à la présence de tumeurs multifocales.

L'ensemble de ces études visant à définir les caractéristiques morphologiques et moléculaires ont pour but d'améliorer la prise en charge du patient en identifiant de nouvelles cibles thérapeutiques. Par exemple, la déficience du mécanisme de réparation des cassures à l'ADN par recombinaison homologue dans les tumeurs *BRCA1* et *BRCA2* a permis d'ouvrir la voie à de nouvelles pistes de traitement avec les inhibiteurs de PARP (Olaparib)^{351, 352} ou les cisplatines. Depuis 2010, l'histologie des tumeurs est également incorporée dans l'algorithme de prédiction BOADICEA (*Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*) calculant la probabilité de porter ou non une mutation dans *BRCA1* ou *BRCA2*. La probabilité pour une femme de moins de 50 ans d'être porteuse *BRCA1* est de 5% si elle développe une tumeur RO+ et de 24% si la tumeur est triple-négative avec des caractéristiques basale-like³⁵³.

Le second objectif de ces travaux est d'identifier des bio-marqueurs tumoraux qui permettraient de discriminer, hors d'un contexte de prédisposition familiale connue, les porteurs d'une mutation dans un gène de prédisposition au cancer du sein afin de lui proposer au patient diagnostiqué avec un cancer du sein et à sa famille, un suivi adapté.

2. Problématique posée dans le cadre des porteurs de variants d'ATM

Peu de données sur la morphologie et sur les caractéristiques moléculaires des tumeurs *ATM* ont été rapportées dans la littérature ; lorsque j'ai commencé ce travail de thèse en 2014, ces données ne permettaient pas de conclure sur l'existence d'une signature tumorale du sein associée aux mutations *ATM*.

La protéine *ATM* étant impliquée dans la réparation des cassures doubles-brins de l'ADN par recombinaison homologue, l'inactivation d'*ATM* dans les tumeurs pourrait se traduire par la perte de grandes régions chromosomiques associées la signature HRD comme montré dans les tumeurs *BRCA1* et *BRCA2*. Par ailleurs, une signature *ATM* pourrait également être caractérisée par des gains et des pertes, des SNVs, des insertions/délétions ou des grands réarrangements sur l'ensemble du génome.

L'objectif de cette deuxième partie de ma thèse était donc de décrire la morphologie et les caractéristiques moléculaires des tumeurs du sein des porteurs d'une ou deux copies altérées du gène afin de mieux comprendre le rôle d'*ATM* dans l'étiologie du cancer et d'identifier des biomarqueurs discriminant ces tumeurs des tumeurs dites sporadiques. De plus, l'identification d'une signature *ATM* pourrait permettre de développer de nouvelles cibles thérapeutiques pour améliorer la prise en charge des patients.

3. Population et matériel étudiés

3.1 Familles A-T

Pour réaliser ces travaux, je me suis appuyée en grande partie sur les ressources biologiques des études CoF-AT et Rétro-AT précédemment décrites (Partie II-2). Lors de la mise en place de CoF-AT une tumorotheque a été établie pour collecter le matériel tumoral des participantes ayant développé un cancer. Au total, 22 tumeurs mammaires provenant de 20 femmes HetAT et de trois enfants atteints d'A-T ont pu être collectées par la PIGE. Afin d'identifier des biomarqueurs spécifiques des tumeurs associées à *ATM*, j'ai exclu des analyses deux tumeurs du sein de femmes HetAT également porteuses hétérozygotes d'une mutation constitutionnelle dans le gène *BRCA1* ou *BRCA2*.

Finalement, 23 tumeurs fixées dans le formol, ou dans le bouin, et enveloppées dans la paraffine (« FFPE » pour *formalin-fixed paraffin-embedded*) ont été collectées. En plus du matériel FFPE, le matériel congelé de 4 tumeurs a également pu être exploité pour ce projet (voir Tableau 1 de l'article rapportant les résultats de cette analyse, p.22-23 de l'article). Parmi les 18 femmes HetAT pour lesquelles le matériel tumoral a été collecté, deux ont été atteintes d'un cancer du sein bilatéral

soit synchrone (T0077) soit asynchrone (T0016).

3.2 Familles prédisposées au cancer du sein

Comme nous l'avons vu précédemment, des mutations d'*ATM* sont identifiées dans des familles à haut risque de cancer du sein et/ou de l'ovaire indépendamment d'un contexte A-T. Des tumeurs de sujets HetAT issus de ces familles ont également été étudiées ici. Ces porteurs ont été identifiés dans les études GENESIS et kConFab portant sur la prédisposition familial au cancer du sein dans le cadre de grands projets de reséquençage.

3.2.1 Etude GENESIS

L'étude GENESIS, pour GENE SISTER, est une étude française de type cas-témoins réalisée entre 2007 et 2013, et visant à identifier de nouveaux gènes de prédisposition au cancer du sein³⁵⁴.

Toutes femmes vues en consultation de génétique dans l'un des centre du Groupe Génétique et Cancer (GGC), atteintes d'un carcinome infiltrant ou d'un carcinome canalaire *in situ*, non porteuses d'une mutation dans les gènes *BRCA1* et *BRCA2* et ayant au moins une sœur atteinte de cancer du sein étaient éligibles à l'inclusion dans l'étude. Les sœurs atteintes des cas index ont également été invitées à participer. La population témoin de GENESIS a été recrutée de deux façons : un premier groupe de témoins est constitué des sœurs non-atteintes de cancer et le deuxième groupe est constitué de femmes non-apparentées (amies, collègues ...) appariées sur l'âge des cas à plus ou moins 3 ans. A l'inclusion, les participantes ont rempli un questionnaire épidémiologique portant sur le mode de vie, l'histoire familiale et sur les données cliniques de la patiente et de ses apparentées au premier et second degré. De plus, pour les cas index, les sœurs atteintes et non-atteintes participantes un échantillon sanguin a été prélevé. Au total, plus de 5000 femmes ont accepté de participer à l'étude, dont la moitié étaient atteintes d'un cancer du sein à l'inclusion dans l'étude.

Dans le cadre d'un projet de reséquençage parallèle massif mené dans GENESIS par notre équipe, en collaboration avec le Centre National de Génotypage (projet France Génomique), la séquence codante d'un panel de 114 gènes de la réparation a été réalisé chez 1205 cas indépendants (cas index ou sœurs atteintes) et 1207 témoins de population. Parmi les variants rares ou absents dans la population témoins, 17 cas portent un variant perte de fonction (variant tronquant) et 25 cas portent une substitution faux-sens prédite comme probablement délétère d'après l'outil Align-GVGD (grade C45, C55 ou C65) et ne présentent pas de variant délétère dans un autre gène connu de prédisposition au cancer du sein.

Onze tumeurs mammaires conservées en FFPE provenant de 11 de ces patientes ont pu être collectées par la PIGE, dont 6 tumeurs de patientes porteuses d'un variant tronquant et 5 tumeurs de

patientes porteuses d'un variant faux-sens classées soit C65 soit C45. (Tableau 1 de l'article rapportant les résultats de cette analyse, p.152-153).

3.2.2 Etude kConFab

La ressource du *Kathleen Cuninghams Foundation Consortium for research into Familial Breast cancer* (kConFab) a été mise en place en 1997 en Australie et en Nouvelle-Zélande afin d'étudier la prédisposition familiale aux cancers du sein et de l'ovaire³⁵⁵. Les données épidémiologiques, médicales et psychologiques sont collectées pour chaque famille participante. De plus, un prélèvement sanguin est réalisé à l'inclusion des participants et les échantillons tumoraux sont collectés afin de caractériser les mutations constitutionnelles dans les gènes de prédisposition au cancer.

Une collaboration avec l'équipe de Georgia Chenevix-Trench, l'une des initiatrices de kConFab, a permis d'utiliser le matériel tumoral (tumeurs conservées en FFPE) provenant de 7 participantes HetAT dont cinq porteuses d'une mutation tronquante d'*ATM* et d'un homme et d'une femmes porteurs de variants faux-sens classés respectivement C65 et C55 (Tableau 1 de l'article rapportant les résultats de cette analyse, p.22-23 de l'article). Aucune des participantes sélectionnées pour ce projet n'est porteuse de la mutation c.7271T>C (p.Val2424Gly) retrouvées dans plusieurs familles australiennes et dont les tumeurs ont déjà été étudiées^{137, 356-358}.

Au total, 36 tumeurs invasives et 5 tumeurs *in situ* de sujets HetAT ou de patients A-T étaient disponibles pour notre étude. L'ADN constitutionnel (sang) de tous les participants était disponible et les comptes rendus anatomopathologiques ont été collectés pour chaque tumeur reçue.

4. Article 2 : Caractéristiques histologiques et moléculaires des tumeurs ATM.

La morphologie de 23 tumeurs de patients de familles A-T ainsi que 18 tumeurs de sujets HetAT a été annotée par les deux pathologistes de référence pour l'étude, le Pr Anne Vincent-Salomon et le Dr Guillaume Bataillon (Département de Biologie des tumeurs – Service de Pathologie, Institut Curie). J'ai ensuite comparé la morphologie de ces tumeurs à celle d'une série clinique témoin (série PICBIM de l'institut Curie).

Au niveau génomique, j'ai recherché les régions de LOH et ai analysé les variations du nombre de copies à partir des données de génotypage de l'ADN tumoral obtenues avec la puce à ADN OncoScan (Affymétrie) sur 23 tumeurs pour lesquelles la qualité et la quantité d'ADN étaient suffisantes. De plus, le séquençage du génome entier de 3 patients HetAT pour lesquels le matériel tumoral congelé était disponible (4 tumeurs en tout) a été réalisé sur la plateforme de séquençage NGS de l'Institut Curie afin d'identifier de potentielles mutations driver.

Les résultats de ces travaux sont décrits dans le manuscrit ci-après soumis pour publication à Cancer Research.

[Click here to view linked References](#)

1 Morphology and genomic hallmarks of breast tumours developed by ATM deleterious 2 variant carriers

3
4
5
6
7 4 Anne-Laure Renault^{1,2,3,4} (anne-laure.renault@curie.fr), Noura Mebirouk^{1,2,3,4}
8
9
10 5 (noura.mebirouk@curie.fr), Laetitia Fuhrmann⁵ (laetitia.fuhrmann@curie.fr), Guillaume Bataillon⁵
11
12 6 (guillaume.bataillon@curie.fr), Eve Cavaciuti^{1,2,3,4} (eve.cavaciuti@curie.fr), Dorothée Le Gal^{1,2,3,4}
13
14 7 (dorothee.legal@curie.fr), Elodie Girard^{1,2,3,4} (Elodie.Girard@curie.fr), Tatiana Popova^{6,2,3}
15
16 8 (tatiana.popova@curie.fr), Philippe La Rosa^{1,2,3,4} (Philippe.La-Rosa@curie.fr), Juana Beauvallet^{1,2,3,4}
17
18 9 (juana.beauvallet@curie.fr), Séverine Eon-Marchais^{1,2,3,4} (Severine.Eon-Marchais@curie.fr), Marie-
19
20 10 Gabrielle Dondon^{1,2,3,4} (marie-gabrielle.dondon@curie.fr), Catherine Dubois d'Enghien⁷
21
22 11 (catherine.dubois-denghien@curie.fr), Anthony Laugé⁷ (anthony.lauge@curie.fr), Walid Chemlali⁸
23
24 12 (walid.chemlali@curie.fr), Virginie Raynal⁹ (Virginie.Raynal@curie.fr), Martine Labbé^{1,2,3,4}
25
26 13 (martine.labbe@gustaveroussy.fr), Ivan Bièche⁸ (ivan.bieche@curie.fr), Sylvain Baulande⁹
27
28 14 (sylvain.baulande@curie.fr), Jacques-Olivier Bay¹⁰ (jobay@chu-clermontferrand.fr), Pascaline
29
30 15 Berthet¹¹ (pberthet@baclesse.unicancer.fr), Olivier Caron¹² (olivier.caron@gustaveroussy.fr), Bruno
31
32 16 Buecher⁷ (bruno.buecher@curie.fr), Laurence Faivre^{13,14} (laurence.faivre@chu-dijon.fr), Marc
33
34 17 Fresnay¹⁵ (marc.frenay@nice.unicancer.fr), Marion Gauthier-Villars⁷ (marion.gauthier-
35
36 18 villars@curie.fr), Paul Gesta¹⁶ (paul.gesta@ch-niort.fr), Nicolas Janin¹⁷ (nicolas.janin@uclouvain.be),
37
38 19 Sophie Lejeune¹⁸ (sophie.lejeune@chru.lille.fr), Christine Maugard^{19,20} (christine.maugard@chru-
39
40 20 strasbourg.fr), Sébastien Moutton²¹ (sebastien.moutton@chu-bordeaux.fr), Laurence Venat-Bouvet²²
41
42 21 (laurence.venat@chu-limoges.fr), Hélène Zattara²³ (Helene.ZATTARA@ap-hm.fr), Jean-Pierre
43
44 22 Fricker²⁴ (Jfricker@strasbourg.fnclcc.fr), Laurence Gladieff²⁵ (gladieff.laurence@claudiusregaud.fr),
45
46 23 Isabelle Coupier^{26,27} (i-coupier@chu-montpellier.fr), CoF-AT^{1,2,3,4}, GENESIS^{1,2,3,4}, kConFab^{28,29}, Georgia
47
48 24 Chenevix-Trench³⁰ (Georgia.Trench@qimrberghofer.edu.au), Janet Hall^{31,32,33*} (janet.hall@inserm.fr),
49
50 25 Anne Vincent-Salomon^{5*} (anne.salomon@curie.fr), Dominique Stoppa-Lyonnet^{7,6,34*}
51
52 26 (Dominique.Stoppa-Lyonnet@curie.fr), Nadine Andrieu^{1,2,3,4*} (nadine.andrieu@curie.fr), Fabienne
53
54
55
56
57
58
59
60
61
62
63
64
65

27 Lesueur^{1,2,3,4§} (fabienne.lesueur@curie.fr)

28

29 *These authors contributed equally

30

31 **Running title:** ATM-associated breast tumour profile.

32

33 **Keywords:** ATM, breast tumour, pathology, genetic instability, OncoScan array, copy number, loss of
34 heterozygosity, genomic signature.

35

36 1- INSERM, U900, Paris, France

37 2- PSL Research University, Paris, France

38 3- Institut Curie, Paris, France

39 4- Mines Paris Tech, Fontainebleau, France

40 5 - Service de Pathologie, Institut Curie, Paris, France

41 6- INSERM, U830, Paris, France

42 7- Service de Génétique, Institut Curie, Paris, France

43 8- Unité de Pharmacogénomique, Institut Curie, Paris France

44 9- Institut Curie Genomics of Excellence (ICGex) Platform, Institut Curie, Paris France

45 10- CHU Clermont-Ferrand, site Estaing, Clermont-Ferrand, France

46 11- Centre François Baclesse, Unité de Pathologie Gynécologique, Caen, France

47 12- Gustave Roussy, Service d'Oncologie génétique, Villejuif, France

48 13- Institut GIMI, CHU de Dijon, Hôpital d'Enfants, Dijon, France

49 14- Oncogénétique, Centre de Lutte contre le Cancer Georges François Leclerc, Dijon, France

50 15- CLCC Antoine Lacassagne, Département d'Hématologie et d'Oncologie Médicale, Nice, France

51 16- Centre Hospitalier Georges-Renon, Service Oncogénétique pour la consultation oncogénétique
52 régionale Poitou-Charentes, Niort, France

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- 53 17- Clinique Universitaire Saint-Luc, Service de Génétique, Brussels, Belgium
- 54 18- Hôpital Jeanne de Flandre, Service de génétique clinique Guy Fontaine, Lille, France
- 55 19- Hôpitaux Universitaires de Strasbourg, UF1422 Oncogénétique moléculaire, Laboratoire de
- 56 diagnostic génétique, Strasbourg, France
- 57 20- Hôpitaux Universitaires de Strasbourg, UF6948 Oncogénétique, Oncogénétique Evaluation
- 58 familiale et suivi, Strasbourg, France
- 59 21- CHU de Bordeaux-GH Pellegrin, Laboratoire Maladies Rares : Génétique et Métabolisme,
- 60 Bordeaux, France
- 61 22- Service d'Oncologie Médicale, Hôpital Universitaire Dupuytren, Limoges, France
- 62 23- Hôpital de la Timone, Département de Génétique, Marseille, France
- 63 24- Centre Paul Strauss, Unité d'Oncogénétique, Strasbourg, France
- 64 25- Institut Claudius Regaud, IUCT Oncopole, Toulouse, France
- 65 26-Hôpital Arnaud de Villeneuve, CHU Montpellier, Service de Génétique médicale et
- 66 Oncogénétique, Montpellier, France
- 67 27- ICM Val d'Aurelle, Unité d'Oncogénétique, Montpellier, France
- 68 28- Research Department, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia
- 69 29- The Sir Peter MacCallum Department of Oncology University of Melbourne, Parkville, Australia
- 70 30- Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute,
- 71 Brisbane, Australia
- 72 31- UMR INSERM 1052, Lyon, France
- 73 32- CNRS 5286, Lyon, France
- 74 33- Centre de Recherche en Cancérologie de Lyon, Lyon, France
- 75 34- Université Paris Descartes, Paris, France
- 76
- 77 ^sTo whom correspondence should be addressed (fabienne.lesueur@curie.fr)
- 78

79 **Abstract**

1
2
3 80 *Background.* The *ataxia telangiectasia-mutated (ATM)* gene is a moderate-risk breast cancer
4
5 81 susceptibility gene; germline loss-of-function variants are found in up to 3% of hereditary breast and
6
7 82 ovarian cancer (HBOC) families who undergo genetic counselling. So far, no clear histopathological
8
9
10 83 and molecular features of breast tumours occurring in *ATM* deleterious variant carriers have been
11
12 84 described, but identification of an *ATM*-associated tumour signature may help patients'
13
14
15 85 management.

16
17 86 *Methods.* To characterize hallmarks of *ATM*-associated tumours, we performed systematic pathology
18
19
20 87 review of tumours from 21 patients from ataxia-telangiectasia families and 18 patients from HBOC
21
22 88 families, and copy number profiling on a subset of 23 tumours. Morphology of *ATM* tumours was
23
24
25 89 compared with those of 599 patients with no *BRCA1* and *BRCA2* mutations from a hospital-based
26
27 90 series, and with data from the Cancer Genome Atlas. Absolute copy number and loss of
28
29 91 heterozygosity (LOH) profiles were obtained from the OncoScan SNP array. In addition we performed
30
31
32 92 whole-genome sequencing on four tumours from *ATM* loss-of-function variant carriers with available
33
34 93 frozen material.

35
36
37 94 *Results.* We found that *ATM*-associated tumours mostly belong to the luminal B subtype, are
38
39 95 tetraploid and show LOH at *ATM* locus at 11q22-23. Unlike tumours in which *BRCA1* or *BRCA2* are
40
41
42 96 inactivated, tumours arising in *ATM* deleterious variant carriers are not associated with increased
43
44 97 large-scale genomic instability as measured by the large-scale state transitions signature. Losses at
45
46 98 13q14.11-q14.3, 17p13.2-p12, 21p11.2-p11.1 and 22q11.23 were observed. Somatic alterations at
47
48
49 99 these loci may therefore represent biomarkers for *ATM* testing and harbour driver mutations in
50
51 100 potentially “druggable” genes that would allow patients to be directed towards tailored therapeutic
52
53
54 101 strategies.

55
56 102 *Conclusions.* Although *ATM* is involved in the DNA damage response *ATM*-associated tumours are
57
58
59 103 distinct from *BRCA1*-associated tumours in terms of morphological characteristics and genomic

104 alterations, and are also distinguishable from sporadic breast tumours, thus opening up the
1
2 105 possibility to identify *ATM* variant carriers outside of an A-T familial context and orientate them
3
4 106 towards effective cancer risk management and therapeutic strategies.
5
6

7 107

10 108 **Background**

11
12 109 Ataxia-telangiectasia (A-T) is a rare autosomal recessive disorder caused by biallelic inactivating
13
14 110 variants in the *ataxia-telangiectasia mutated (ATM)* gene. The phenotype is characterized by
15
16
17 111 progressive neuronal degeneration, immunological deficiency, genetic instability, hypersensitivity to
18
19 112 ionizing radiation and agents that cause DNA double strand breaks, as well as a predisposition to
20
21
22 113 malignancies, particularly lymphoid tumours [1-3]. Epidemiological studies on A-T families showed
23
24 114 that heterozygous *ATM* deleterious variant carriers (hereafter called “HetAT” subjects) are also at
25
26 115 increased risk of other cancer types [4-6], notably of breast cancer (BC) in female relatives [7, 8]. It is
27
28
29 116 estimated that 0.5 to 1% of the general population carry an *ATM* deleterious variant, and studies
30
31 117 conducted in hereditary breast and ovarian cancer (HBOC) families or early onset BC cases described
32
33 118 *ATM* as a moderate BC susceptibility gene, with a 3-fold increase in risk of developing BC for HetAT
34
35
36 119 subjects as compared to non-carriers [9, 10]. Consequently, *ATM* is now included in nearly all
37
38 120 multigene panels used for HBOC genetic testing that include, in addition to *BRCA1* and *BRCA2*, other
39
40
41 121 moderate- to high-risk genes coding for tumour suppressor proteins acting in critical processes of
42
43 122 DNA repair pathways [11]. However, results for *ATM* testing are usually not issued to patients
44
45 123 because of the imprecise absolute risk estimates and a lack of management recommendations for
46
47
48 124 *ATM* variant carriers and their relatives [11]. If a consensus was made to define *ATM* as a gene with
49
50 125 clinical utility, specific pathological and genomic features associated with *ATM* inactivation in
51
52 126 tumours could help identify patients for whom a genetic test for *ATM* may be offered, and therefore
53
54
55 127 direct those patients toward effective cancer risk management and therapeutic strategies.
56

57 128 So far, no clear histopathological and molecular features have been described for *ATM*-associated
58
59 129 breast tumours, and well-documented *ATM* tumour series are very limited. Three Australian studies
60
61
62
63
64
65

130 on familial BC investigated breast tumours developed by HetAT patients carrying either a truncating
1
2 131 variant or a missense variant (essentially the c.7271T>G; p.Val2424Gly variant) [12, 13]. The first two
3
4 132 studies carried out on 21 and 35 tumours respectively suggested that histologically, breast tumours
5
6 133 from HetAT subjects do not resemble the tumours from *BRCA1* mutation carriers [12], and no
7
8 134 difference was observed between the histological grade of *ATM* tumours and a series of age-
9
10 135 matched control tumours [13]. The third study focusing on six tumours from carriers of the
11
12 136 c.7271T>G variant revealed that all tumours presented with the luminal A or B molecular subtypes
13
14 137 [14]. Finally, no consistent pattern of loss of the normal allele was reported in an Australian series
15
16 138 (*N*=17) and in a French series (*N*=16) of tumours from carriers of putative BC-associated *ATM* variants
17
18 139 [13, 15].
19
20
21
22
23

24 140 The purpose of the present study was to determine whether breast tumours developed by *ATM*
25
26 141 variant carriers show distinctive histopathological and genomic features as compared to “sporadic”
27
28 142 tumours, and also whether they resemble the breast tumours described in carriers of a deleterious
29
30 143 variant in other known BC susceptibility genes, in particular the *BRCA1*-associated phenotype [16]. To
31
32 144 this end, we conducted pathology review of a series of tumours composed of three breast tumours
33
34 145 from three A-T patients, 20 tumours from 18 HetAT subjects from A-T families, and 18 tumours from
35
36 146 18 HetAT subjects from HBOC families who were non-carriers of other known high-risk variants. We
37
38 147 also performed SNP array genomic profiling to assess somatic loss of heterozygosity (LOH) at the
39
40 148 *ATM* locus and to investigate absolute copy number and LOH profiles at the genome-wide level.
41
42 149 Finally, to complete the repertoire of somatic alterations of *ATM*-associated breast tumours we
43
44 150 performed whole-genome sequencing on four tumours from HetAT patients with available frozen
45
46 151 material.
47
48
49
50
51
52

53 152

54

55 153 **Patients, Material and Methods**

56

57 154 *Patients and tumour material*

58

59

60

61

62

63

64

65

155 Breast tumour samples were selected from carriers of a deleterious *ATM* variant from four different
1
2 156 research resources: the French retrospective study on A-T families (Retro-AT) [17], the French
3
4 157 prospective cohort on women related to an A-T child (CoF-AT) [8], the GENESIS study [18] and the
5
6
7 158 Kathleen Cuninghame Foundation Consortium for research into Familial Breast Cancer (kConFab)
8
9 159 study [19].

11 160 Briefly, Retro-AT [17] was carried out between 1994 and 1997 to assess cancer risk in A-T families
12
13
14 161 living in France. Thirty-four A-T families were identified during this period and 27 of them were
15
16 162 subsequently included in the CoF-AT prospective study. CoF-AT is an ongoing national prospective
17
18
19 163 cohort on A-T families, which was initiated in 2003 to investigate environmental and genetic risk
20
21 164 factors for BC in heterozygous HetAT and non-HetAT women. All women aged 18 and over are
22
23 165 eligible to participate in the study. At inclusion, participants provide a blood sample to determine
24
25
26 166 whether they carry one of the *ATM* inactivating variant identified in the A-T child of the family. In
27
28 167 June 2017, 415 women (213 HetAT and 202 non-HetAT) belonging to 105 A-T families had been
29
30
31 168 enrolled in CoF-AT, and 37 study participants from Retro-AT or CoF-AT had developed BC, including
32
33 169 23 HetAT women, 11 non-HetAT women and 3 A-T patients (2 females and 1 male). Breast tumour
34
35 170 material from the 3 A-T patients and from 18 HetAT subjects could be retrieved for the present study
36
37
38 171 (Table 1).

40 172 GENESIS is a national study on HBOC families identified through the French family cancer clinics [18].
41
42 173 Index cases are women diagnosed with invasive mammary carcinoma or *in situ* ductal carcinoma,
43
44
45 174 having at least one sister affected with BC, and negatively tested for a pathogenic variant in *BRCA1*
46
47 175 and *BRCA2*. *ATM* carriers of a truncating variant (TV) or of a rare likely deleterious missense
48
49
50 176 substitution (MS) were identified during the course of a large-scale case-control mutation-screening
51
52 177 study (Girard *et al.* manuscript in preparation). Tumour material from 11 of them was assessed in the
53
54 178 present study (Table 1).

56 179 In addition, we investigated tumours from seven HetAT subjects enrolled in the Australian Kathleen
57
58
59 180 Cuninghame Consortium for Research into Familial Breast Cancer (kConFab) study [19] (Table 1).

181

1
2 182 *Selection of ATM variant carriers*

3
4
5 183 Individuals included in the study were either homozygous, compound heterozygous or heterozygous
6
7 184 carriers of a variant considered pathogenic for A-T disease. We also selected HetAT BC patients from
8
9 185 HBOC families carrying a TV that had not necessarily been reported in A-T families, as well as carriers
10
11 186 of a rare likely deleterious MS classified as C65, C55 or C45 according to the Align-GVGD tool as
12
13 187 previously described [10, 20]. Carriers of the p.Val2424Gly variant identified in kConFab were not
14
15 188 included in this study since tumour characteristics of carriers of this variant had been already
16
17 189 investigated [12-14]. In total, 41 tumour tissues were available from 3 A-T patients and 36 HetAT
18
19 190 patients for histopathological review (Table 1).
20
21
22
23

24 191

25
26 192 *Pathology review*

27
28
29 193 The hematoxylin, eosin and safran-stained breast tumour tissue (HES) was reviewed and scored for
30
31 194 morphology features and grade by two pathologists (AV-S and GB) using the modified system of
32
33 195 Elson-Ellis [21]. The World Health Organization (WHO) classification of tumours of the breast was
34
35 196 used to determine the histological subtypes of *ATM*-associated tumours and the TNM stage
36
37 197 according to tumour size (T), nodal infiltration (N) and metastasis status (M) [22]. ER, PR and HER2
38
39 198 status, as well as expression of proliferating marker Ki67 were obtained from histopathology reports
40
41 199 held by diagnostic laboratories. When incomplete, hormonal statuses were determined by
42
43 200 immunohistochemical (IHC) staining at Institut Curie. Tumours were considered HER2+ if they were
44
45 201 scored 3+ by IHC or for tumours scored 2+ by IHC, if the fluorescence in situ hybridization showed an
46
47 202 HER2 gene amplification. Tumours were classified using IHC data according to the St Gallen
48
49 203 molecular subtypes as follows: triple-negative (ER-, PR- and HER2-), HER2-overexpressing (ER-, PR-
50
51 204 and HER2+), luminal A (ER+, PR+/-, HER2- and Ki67<20%), luminal B (ER+, PR+/-, HER2- and
52
53 205 Ki67≥20%), and luminal B/HER2+ (ER+, PR+/-, HER2+ and Ki67≥20%) [23].
54
55
56
57
58
59

60 206 Morphology features of *ATM*-associated breast tumours were compared to the series of breast
61
62
63
64
65

207 tumours from patients who received a surgery at Institut Curie between 2005 and 2006, named the
1
2 208 PICBIM series (from “*programme incitatif et collaboratif - Cancer du sein: invasion et motilité*”). None
3
4 209 of the PICBIM patients received neo-adjuvant treatment. Patients who had developed a previous
5
6 210 cancer at any site were excluded, as well as known *BRCA1* or *BRCA2* mutation carriers. In total, 516
7
8 211 cases diagnosed with invasive carcinoma and 83 cases diagnosed with *in situ* carcinoma served as
9
10 212 controls.

11
12
13
14 213

15 16 214 *DNA preparation and confirmation of the familial ATM deleterious variant*

17
18
19 215 Tumour DNA was extracted from tumour-enriched areas (with at least 50% tumour content when
20
21 216 possible) delimited from the most representative HES slides for the 35 tumours for which formalin-
22
23 217 fixed paraffin-embedded (FFPE) material was available (Table 1). The relevant areas were macro-
24
25 218 dissected from four 10µm sections and DNA was purified using the NucleoSpin Tissue protocol
26
27 219 according to the manufacturer’s instructions (Macherey-Nagel, Düren). DNA quantity and quality was
28
29 220 assessed using QuBit fluorometry (Life Technologies) and SYBR-Green-based qPCR assay (Promega).
30
31 221 Of the 35 available FFPE tumour DNA samples, sufficient quantity and quality to perform subsequent
32
33 222 molecular analyses were obtained for 23 of them. Matched blood DNA was extracted with the
34
35 223 Quickgene 610-L automated system (Fujifilm) according to the manufacturer’s instructions. The
36
37 224 presence of the familial *ATM* deleterious variant was confirmed in all analysable blood and tumour
38
39 225 DNA samples by Sanger targeted sequencing on the ABI3500xL DNA analyser (Applied Biosystems,
40
41 226 Forest City, CA).

42
43
44
45 227

46 47 228 *Copy Number Variation analysis*

48
49
50 229 CNVs analysis using the OncoScan SNP array (Affymetrix) was performed on the 23 FFPE *ATM*-
51
52 230 associated tumours, including two tumours from two A-T patients (Table 1). Data were analysed with
53
54 231 the Genome Alteration Print (GAP) method that takes into account both ploidy and large-scale
55
56 232 genomic rearrangements [16, 24]. Copy number was ranged from 0 to 8 copies and all segments

233 exceeding 8 copies were ascribed 8-copy status. Chromosome number was estimated by the sum of
1
2 234 the copy numbers detected at the pericentric regions. Output processing files from GAP were
3
4 235 analysed using the VAMP in-house software [25] to define the boundaries of regions recurrently
5
6
7 236 altered in *ATM*-mutated tumours. Copy loss and gain for near-diploid tumours were called for the
8
9 237 segments with ≤ 1 and ≥ 4 copies, respectively. Copy loss and gain for near-tetraploid tumours were
10
11 238 called for the segments with ≤ 2 and ≥ 6 copies, respectively. LOH status was ascribed to regions
12
13
14 239 having monoallelic content irrespective of copy number. LOH associated with copy loss was referred
15
16 240 as LOH/loss. Breakpoints (changes in the copy number or major allele counts within chromosomes) in
17
18
19 241 each genomic profile were characterized based on the resulting absolute copy number profile and
20
21 242 after filtering for region with less than 50 SNP variations. Recurrent alterations (CNV, LOH) among
22
23 243 the cohort were obtained using the CNTools R package (v1.24.0) and a homemade script. In order to
24
25
26 244 find inherent grouping structure, a hierarchical clustering was performed using the alteration status
27
28 245 (absence/ presence of an alteration) per segment using the Jaccard distance and the Ward linkage
29
30
31 246 function, available in the vegan R package (v2.3-3).

32
33 247

34 35 248 *Validation of LOH status at ATM locus using microsatellites*

36
37
38 249 Tumour and matched blood DNAs were evaluated on a subset of patients by a PCR-based LOH assay
39
40 250 using four fluorescence-labelled microsatellite markers (namely D11S1113, D11S1819, D11S2179 and
41
42
43 251 D11S1778) spanning a 14.4Mb region encompassing the *MRE11A* and *ATM* genes. Capillary
44
45 252 electrophoresis was performed on the ABI3500xL DNA analyser (Applied Biosystems). Raw
46
47
48 253 electrophoresis data were analysed with GeneMarker software version 1.3 (Softgenetics, LLC, State
49
50 254 College) to assess allele ratios. We considered LOH at *ATM* locus when allele ratio fell below 50% in
51
52 255 the tumour DNA sample.

53
54 256

55 56 57 257 *Whole-genome sequencing*

58
59 258 Whole genome sequencing (WGS) was performed on four tumour/normal DNA pairs from three
60
61
62
63
64
65

259 patients for whom frozen tumour tissue was available (Table 1). Paired-end libraries were prepared
1
2 260 from 2µg DNA using the TruSeq DNA PCR-Free LT Sample Preparation kit (Illumina, San Diego, CA,
3
4 261 USA) and were sequenced on an Illumina HiSeq2500 instrument (Illumina). Tumour DNAs were
5
6
7 262 sequenced at a higher depth of coverage (100X) than their germline counterpart (30X). Sequencing
8
9 263 reads were mapped to the reference genome (assembly hg19) using BWA (v0.7.5a)[26]. Regions of
10
11 264 CNV and LOH were identified using Facets algorithm (v0.5.6)[27], and single nucleotide variations
12
13
14 265 (SNVs) and indels were called using VarScan2[28]. Somatic variants were filtered and annotated using
15
16 266 an in-house pipeline.

17
18
19 267

21 268 *Statistical analysis*

23
24 269 Statistical analyses were performed using STATA, version 14.1 (StataCorp, College Station, TX). Two-
25
26 270 tailed tests with 5% significance level were used throughout. Logistic regressions were used to assess
27
28 271 the level of association between the presence of an *ATM* variant and various features of interest
29
30
31 272 when comparing the *ATM* series with the PICBIM series. A Fisher's exact test (FET) was used to assess
32
33 273 molecular subtypes difference between *ATM*-associated tumours and breast tumours from sporadic
34
35
36 274 cases from The Cancer Genome Atlas (TCGA) [29] and from the Norwegian series [30].

37
38 275

39
40
41 276

43 277 **Results**

46 278 *Histopathological features associated with ATM mutation status*

48
49 279 Clinicopathological and immunohistochemical features were evaluated on three breast tumours
50
51
52 280 from three A-T patients and 38 tumours from 36 HetAT patients. This tumour series was compared to
53
54 281 BC cases enrolled in the PICBIM program of Institut Curie. An overview of the features examined in
55
56 282 both series is presented in Table 2.

57
58
59
60
61
62
63
64
65

283 Among the 41 reviewed *ATM*-associated breast tumours, 36 were invasive carcinomas and five were
1
2 284 *in situ* carcinomas. Overall, subjects with invasive carcinoma and subjects with *in situ* carcinoma from
3
4 285 the *ATM* series tend to be diagnosed at a younger age than subjects from the PICBIM series (mean
5
6
7 286 age: 52.4 vs. 56.2, $P=0.08$ for invasive carcinomas; 45.5 vs. 54.1, $P=0.07$ for *in situ* carcinomas). This
8
9 287 can be explained by the fact that women related to an A-T child or belonging to a HBOC family are
10
11 288 more likely to benefit from early detection of the disease due their higher risk of developing breast
12
13
14 289 cancer than the general population. We also compared mean age at diagnosis in patients having
15
16 290 developed invasive breast carcinoma between the studies and observed no difference (CoF-AT/
17
18 291 Retro-AT: 51.9; GENESIS: 55.9; kConFab: 50.6, $P=0.90$).

20
21 292 Invasive breast carcinomas developed by *ATM* variant carriers were mostly ductal carcinomas (86%)
22
23 293 with an intermediate to high grade (II-III) which was the same distribution of the histological type
24
25
26 294 and grade as found in the invasive tumours from the PICBIM series. With respect to the
27
28 295 immunohistochemistry of tumours arising in *ATM* variant carriers, 97% of *ATM* tumours were ER+,
29
30 296 which was significantly higher than the proportion of ER+ tumour in PICBIM series (59%, $P=0.004$).
31
32
33 297 Low to moderate lymphocytic infiltration was observed in *ATM*-associated tumours (data not shown)
34
35 298 but this information was not available in the PICBIM, so no comparison could be performed.

36
37 299 Molecular subtypes could be determined for 28 out of 36 *ATM*-associated invasive breast tumours.
38
39
40 300 *ATM*-associated breast tumours mostly harboured luminal B (46%) and luminal A (36%) profiles, and
41
42 301 the distribution of the molecular subtypes differed significantly from that of the PICBIM series. In
43
44
45 302 particular the luminal B and luminal B/HER2+ subtypes were more represented among tumours
46
47 303 developed by *ATM* variant carriers ($P=0.009$ and $P=0.005$ respectively) (Table 2). As the PICBIM series
48
49
50 304 might not reflect the distribution of the molecular subtypes of invasive breast tumours in the general
51
52 305 population, we also compared the *ATM* series to a series of 1,423 primary breast tumours from a
53
54 306 Norwegian population-based survey of women born between 1886 and 1977 [30], and to 501
55
56
57 307 invasive breast tumours characterized with the PAM50 test [31, 32] available in TCGA, after exclusion
58
59 308 of carriers of a TM in *BRCA1*, *BRCA2* and *ATM* [29]. We found that the proportion of luminal B
60
61
62
63
64
65

309 tumours was also significantly higher in *ATM*-associated tumours than in the Norwegian study
1
2 310 ($P_{\text{FET}}=0.03$) and in the TCGA series ($P_{\text{FET}}=0.02$), while the prevalence of luminal B/HER2+ tumours and
3
4 311 of triple-negative breast tumours were similar in the two latter series to those observed in *ATM*-
5
6 312 associated tumours (Figure 1).
7
8

9 313 We next performed two sensitivity tests. First, analyses comparing clinical and histological features
10
11 314 of *ATM*-associated tumours to those of the PICBIM series were repeated after exclusion of four *ATM*-
12
13 315 associated invasive breast tumours, three of which were second primary tumours (T0016-L, T0173-R
14
15 316 and T0179-L) whose morphology and histology might have been affected by treatment of the first
16
17 317 primary BC. The fifth tumour (T0077-L) was from a patient with synchronous bilateral tumours
18
19 318 (T0077-R and T0077-L); we excluded randomly one of the two tumours in order to take into account
20
21 319 only one tumour per patient in the analysis. Results remained unchanged (Table 2).
22
23
24

25 320 Second, we repeated the analyses after exclusion of five invasive tumours developed by carriers of a
26
27 321 missense variant not being reported so far as pathogenic for A-T, namely c.1464G>T (p.Trp488Cys),
28
29 322 c.5750G>C (p.Arg1917Thr), c.8158G>C (p.Asp2720His), c.8614C>A (p.His2872Asn) and c.9008A>T
30
31 323 (p.Asn3003Ile), in order to avoid possible misclassification of the deleterious effect of the variant
32
33 324 based on *in silico* prediction only. Again, results remained unchanged (Table 2).
34
35
36

37 325 We also compared features of the five *in situ* carcinomas (two ER+, two ER- and one with
38
39 326 undetermined ER status) of the *ATM* series with those of the 83 *in situ* carcinomas from the PICBIM
40
41 327 series, and observed no difference in nuclear grades, tumour size and hormonal status between the
42
43 328 two groups of tumours. However due to low number of *in situ* carcinomas observed in *ATM* variant
44
45 329 carriers in our study it is not possible at this stage to draw any conclusions about the characteristics
46
47 330 of the *in situ* tumours developed by *ATM* variant carriers.
48
49
50

51 331
52
53
54 332 *Genome-wide copy number and LOH profiles of ATM breast tumours*
55

56
57 333 High-quality SNP array data were obtained for 23 FFPE breast tumours; two tumours were from A-T
58
59 334 patients and 21 tumours were from HetAT subjects.
60
61
62
63
64
65

335 Tumour ploidy inferred from the absolute segmental copy number profiles and genotype status by
1
2 336 the GAP method [24] identified 16 out of 23 (70%) near-tetraploid tumours and 7 out of 23 (30%)
3
4 337 near-diploid tumours. Copy number variations and regions of LOH were subsequently determined by
5
6
7 338 taking into account the ploidy of each tumour. No evidence of the homologous recombination
8
9 339 deficiency (HRD) as measured by large-scale state transitions genomic signature [16, 33] was
10
11 340 observed among the *ATM*-associated tumours (Figure 2A).
12
13
14 341 Since “two-hit” inactivation of the causative gene is regarded as a principal feature of molecular
15
16 342 pathogenesis of most hereditary tumours, we next examined the LOH status of *ATM*-associated
17
18 343 tumours at 11q22-23 in the tumours from HetAT patients. LOH was found in 14 out of the 21
19
20
21 344 tumours (67%) developed by HetAT patients. Microsatellite analysis performed on a subset of 14
22
23 345 tumours confirms OncoScan results except for one tumour (T0005). The one exception may be due
24
25
26 346 to differences in sensitivity of the two methods. In the subsequent analyses, OncoScan results were
27
28 347 considered for this tumour. Sanger sequencing of the tumour/blood DNA pairs suggested that the
29
30
31 348 *ATM* wild type allele was lost in all tumours that underwent LOH at this locus (data not shown).
32
33 349 Genome-wide profiling of the 23 *ATM*-associated tumours revealed multiple copy number
34
35 350 aberrations including those previously reported in breast tumours, such as losses at 8p and gains at
36
37
38 351 8q [29] occurring in 50% and 70% of *ATM*-associated tumours, respectively (Figures 2A and 2B). Copy
39
40 352 number losses at 16q, 17p and 22q, which are known features of breast tumours of luminal A and B
41
42 353 subtypes [29, 34] were also seen in 70% of *ATM*-associated tumours (Figure 2B). In addition, 70% of
43
44
45 354 *ATM*-associated tumours showed copy number losses at 13q14.11-q14.3, 17p13.2-p12 and 21p11.2-
46
47 355 p11.1 (Figure 2B). The 13q14.11-q14.3 locus is 9.6Mb long and contains 90 genes including *LCP1*
48
49
50 356 (*Lymphocyte cytosolic protein 1*) and *RB1* (*RB transcriptional co-repressor 1*) (Table 3). This region is
51
52 357 included in the 13q12.3-q21 locus identified in high-grade luminal BRCA2-associated tumours which
53
54 358 was described by Pecuchet et al [35]. The 17p13.2-p12 contains 166 cancer-related genes including
55
56
57 359 *TP53* and *MAP2K4* (Table 3). The 21p11.2-p11.1 locus is 1.2Mb long and only contains the PTEN-
58
59 360 related tyrosine phosphatase gene *TPTE*, the pseudogene *TEKT4P2*, and four miRNA (MIR3648-1,
60
61
62
63
64
65

361 MIR3648-2, MIR3687-1 and MIR3687-2) (Figure 2B and Table 3). The complete list of genes located in
1
2 362 segmental losses observed in at least 70% of *ATM*-associated breast tumours is provided in
3
4 363 Supplementary Table S1.
5
6
7 364 When restricting the analysis to the 16 tumours in which biallelic inactivation of *ATM* was
8
9 365 demonstrated (*i.e.* breast tumours from A-T patients and breast tumours from HetAT patients
10
11 366 showing LOH at *ATM* locus), we found that copy number losses at 8p, 11q, 13q and 22q
12
13 367 corresponded to longer chromosome segments as compared to the ones described in the 23 *ATM*-
14
15 368 associated tumours. The lost segment at 21p11 was the same as the one initially described (Figure 2C
16
17 369 and Table 3).
18
19
20

21 370 We also performed unsupervised hierarchical clustering analyses of CNVs data. This analysis did not
22
23 371 allow separating *ATM*-associated tumours according to molecular subtype, LOH status at 11q22
24
25 372 (*ATM* locus), type of inherited variant (TV vs. MS) or origin of HetAT patients (A-T families or HBOC
26
27 373 families) (Figure 2B). Interestingly, the synchronous bilateral tumours from the HetAT patient T0077
28
29 374 showed similar CNV profiles, while tumours from the two A-T patients showed quite distinctive
30
31 375 features (Figures 2B and 2C).
32
33
34

35 376

38 377 *Comparison with publicly available data*

39
40 378 We used the publicly available data from TCGA [29] accessible through cBioportal [36] to investigate
41
42 379 whether genes listed in Table 3 were specifically lost in *ATM*-associated tumours. A total of 745 TCGA
43
44 380 tumour samples with available CNV data were used. Those tumours were from patients having
45
46 381 developed invasive primary breast tumours and not carrying a deleterious variant in *ATM*, *BRCA1* or
47
48 382 *BRCA2*. In addition to LOH at the *ATM* locus, which was more frequently observed in *ATM*-associated
49
50 383 tumours (67%) than in the TCGA “sporadic” tumours (40.1%) ($P=0.02$), several genes at other loci
51
52 384 appeared more frequently lost in *ATM*-associated tumours, including *TPTE* (21p11.2-p11.1), *GSTT1*,
53
54 385 *GSTTP1* and *GSTTP2* (22q11.23), as well as *LCP1*, *RB1* (13q14), *YWHAE*, *USP6*, *RABEP1* and *MAP2K4*
55
56 386 (17p13.3-p12) (Supplementary Table S2).
57
58
59
60
61
62
63
64
65

387

1

2 388 *Deep whole-genome sequencing of ATM-associated tumours*

3

4

5 389 Whole genome massively parallel sequencing of four *ATM*-associated breast tumours (T0001-L,

6

7 390 T0015-L, T0077-L, T0077-R) and their respective germline DNAs, was used to characterize the genetic

8

9 391 landscape of *ATM*-associated tumours at base-pair resolution. Tumour DNAs were sequenced at a

10

11 392 mean depth of coverage of 97X (range 82X-104X) and paired blood DNA were sequenced at 36X

12

13 393 (range 35X-37X). CNV patterns obtained from WGS data for frozen tumours T0015-L, T0077-L and

14

15 394 T0077-R were compared to CNV pattern obtained in the OncoScan analysis in the corresponding

16

17 395 FFPE tumours. Loss/LOH were confirmed by the whole-genome analysis for loci 8p21.3, 11q21-q24.2,

18

19 396 13q13.3-q32.3, 16q22.1 and 17p13.3, while discordant results were obtained at locus 21p11.2-p11.1

20

21 397 for one tumour, and at locus 22p11.23 for two tumours (Figure 3A). Divergent ploidy estimations

22

23 398 between the WGS analysis (ploidy=3.5) and the OncoScan analysis (ploidy=4), or the use of different

24

25 399 tumour sections to prepare tumour DNA may explain these discrepancies. No OncoScan data were

26

27 400 available for tumour T0001 but the CNV profile obtained from the WGS data showed LOH at 11q21-

28

29 401 q24.2 (containing *ATM*) and also the losses/LOH at 13q13.3-q32.3, 17p13.3 and 22q12.3-q13.31.

30

31 402 In addition we found that the four tumour genomes shared a region of copy number loss/LOH at

32

33 403 6q23.3-q27, which contains *ESR1* encoding the oestrogen receptor, as well as a region of copy

34

35 404 number loss at 19p13.3-p13.2 measuring 7.9Mb (Figure 3A) and containing 256 genes (Table 3).

36

37 405 Going back to the OncoScan data, we found that these two latter regions were indeed altered but in

38

39 406 less than 40% of the analysed FFPE tumour genomes.

40

41 407 Based on the analysis of high-confidence single nucleotide variants (SNVs) identified in each *ATM*-

42

43 408 associated tumour genome, we next looked for potential driver mutations. Post-filtering, 51,161

44

45 409 SNVs were identified, 1004 of them were shared by two tumours, and 29 were shared by three

46

47 410 tumours (Figure 3B). Only 794 SNVs were shared by the synchronous bilateral tumours T0077-L and

48

49 411 T0077-R (Figure 3B). When analyses were restricted to the coding part of the genome (exome) no

50

51 412 genes were found to be altered in all four tumours, nor in the two tumours from patient T0077

52

53

54

55

56

57

58

413 (Figure 3C). Six genes were found to be altered in two tumours: *MYO1A*, *DNAH11*, *SH2D5*, *ATM*,
1
2 414 *MUC4* and *ROS1* (Figure 3C). However only *DNAH11* variants (c.7134+1G>A and c.9255_9257del) and
3
4 415 the nonsense variant in *MUC4* (c.11207C>G) are likely to have a deleterious impact on the gene
5
6 416 product function and therefore might represent candidate driver mutations. The two *ATM* somatic
7
8 417 variants identified in tumours T0015 and T0077-L were predicted as benign variants according to
9
10
11 418 Align-GVGD prediction tool (Figure 3C).
12
13

14 419

15

16

17 420 **Discussion**

18

19

20 421 This exploratory study investigating both the histological and the molecular features of breast
21
22 422 tumours developed by patients having inherited one or two mutated copies of *ATM* describes, to our
23
24
25 423 knowledge, the largest series of *ATM*-associated tumours reported to date. One asset of the study
26
27 424 design is that the vast majority of patients included in the study carried a loss-of-function or
28
29 425 missense variant that had been identified in an A-T family, hence avoiding introducing noise in the
30
31
32 426 analysis that would be due to misclassification of an *ATM* variant. Moreover, all *ATM*-associated
33
34 427 breast tumours and the control series were blind reviewed by trained reference pathologists of
35
36
37 428 Institut Curie (AVS and GB), thus ensuring unbiased scoring of the morphological features.

38

39 429 The study revealed that most *ATM*-associated breast tumours are luminal B or luminal B/HER2+
40
41 430 tumours, which is in accordance with a recent case-control study showing that *ATM* truncating
42
43
44 431 variant carriers are at increased risk of developing ER+ breast tumours [37]. Moreover tetraploidy,
45
46 432 loss of the wild type allele at *ATM* locus, and copy number loss/LOH at loci 13q14.11-q14.3, 17p13.2-
47
48
49 433 p12, 21p11.2-p11.1 and 22q11.23 are hallmarks of breast tumours developed by *ATM* variant
50
51 434 carriers.

52

53 435 In comparison with breast tumours associated with other BC susceptibility genes, we thus confirmed
54
55
56 436 previous observation showing that *ATM*-associated tumours do not resemble *BRCA1* tumours [12,
57
58 437 13] or *PALB2* tumours, which are also predominantly triple-negative tumours [38, 39]. Like *BRCA2*

59

60

61

62

63

64

65

438 and *CHEK2* tumours, *ATM* tumours are mostly luminal tumours [40-42] and do not show a particular
1
2 439 histological subtype as observed in *BRCA1* (medullary) [43], *BRCA2* (lobular) [43], *CDH1* (lobular) [44],
3
4 440 and *PTEN* tumours (apocrine) [45].
5
6
7 441 The absence of histological resemblance between *BRCA1* and *ATM*-associated tumours was reflected
8
9 442 at molecular level. Indeed, *ATM*-associated tumours do not show the HRD signature characterized by
10
11 443 large-scale state transitions [16, 46], suggesting that tumorigenesis in *BRCA1* variant carriers and
12
13 444 *ATM* variant carriers occurs by different mechanisms. *ATM*-associated tumours also differ from
14
15 445 luminal *BRCA2* tumours, which may display the HRD signature [33]. Our results are in accordance
16
17 446 with recently published results on tumour-derived genome sequences from 7 breast cancer patients
18
19 447 from TCGA carrying a TV [47]. Therefore, one cannot predict sensitivity of HetAT subjects to cisplatin
20
21 448 and/or PARP inhibitors.
22
23
24
25
26 449 Interestingly, tumour genomic profile revealed that about 70% of *ATM*-associated breast tumours
27
28 450 are tetraploid. Polyploidy can be triggered by cell fusion, endoreplication or abortive cell cycle [48].
29
30 451 *ATM* being required for three cell cycle checkpoint controls: G1/S border, S phase and G2/M
31
32 452 checkpoints after DNA double-strand breaks, the emergence of polyploidy could be due to cell cycle
33
34 453 checkpoint defects linked to inactivation of *ATM* in breast tumours. Of note, tetraploidy was also
35
36 454 reported in *BRCA2* tumours associated with luminal molecular subtype and loss of the normal allele
37
38 455 [49], although this result was not confirmed when investigating *BRCA2* CNV profile with SNP array
39
40 456 and GAP methods [35].
41
42
43
44
45 457 We found that LOH at *ATM* locus was more frequent in tumours from HetAT subjects (67%) than in
46
47 458 “sporadic” tumours from TCGA (40%) and from previous studies investigating LOH in tumours from
48
49 459 sporadic breast cancer cases (20%-40%) [50, 51]. This observation is in accordance with Knudson’s
50
51 460 “two-hit” model, where the second allele of the tumour suppressor gene would be an early event in
52
53 461 the oncogenic process of hereditary BC. Similar results were found in *BRCA1* and *BRCA2* tumours
54
55 462 [46]. With regard to *ATM*-associated tumours, one cannot exclude that biallelic inactivation of *ATM*
56
57 463 occurs through promoter methylation of the gene in the seven tumours not showing LOH at *ATM*
58
59
60
61
62
63
64
65

464 locus or through point or small size sequence variation (Figure 1C). Another possible explanation for
1
2 465 carriers of a deleterious missense variant would be that such alterations might have a dominant
3
4 466 negative effect, and therefore do not require inactivation of the second allele to impact gene product
5
6
7 467 function. Finally, as previously reported, we did not observe a clear pattern of the biallelic
8
9 468 inactivation of *ATM* according to variant type (truncating variants vs. deleterious or likely deleterious
10
11 469 missense variants) [13].
12
13

14 470 In *ATM*-associated tumours, the cumulative profile of copy number losses, gains and regions in LOH
15
16 471 revealed several genomic regions frequently altered in breast tumours, and in particular in luminal
17
18 472 tumours, which was consistent with the molecular subtypes defined by IHC staining in our *ATM*
19
20
21 473 series. Nevertheless, when comparing to TCGA sporadic tumours, the following copy number losses
22
23 474 or regions of LOH appeared to be specific to *ATM*-associated tumours: 13q14.11-q14.3 (*LHFP*,
24
25 475 *FOXO1*, *LCP1*, *RB1*), 21p11.2-p11.1 (*TPTE*, *TEKT4P2*, *MIR3648-1*, *MIR3648-2*, *MIR3687-1*, *MIR3687-2*)
26
27 476 and 22q11.23 (*GSTT1*, *GSTTP1*, *GSTTP2*). Interestingly, loss of *RB1* was associated with a poor
28
29
30 477 response to hormonotherapy [52] and *LCP1* was considered as biomarker of advance tumour stage
31
32
33 478 and tumour severity [53]. Unfortunately in our study protein expression analysis could not be
34
35 479 performed to validate the diminution of expression of these genes in *ATM*-associated tumours due
36
37 480 to limited material.
38
39

40 481 To complete the repertoire of somatic alterations in *ATM*-associated tumours we performed WGS on
41
42 482 the four frozen tumours available. Only four deleterious variants located in the two genes *DNAH11*
43
44 483 and *MUC4* were identified in two tumours. Little is known on the role of these two genes in
45
46
47 484 tumorigenesis. However, the diminution of *MUC4* expression has been associated with tumour
48
49
50 485 progression, and with an increase infiltration of immune CD8+T and NK cells [53]. Remarkably, no
51
52 486 mutation in *TP53* and *PIK3CA* was detected in any of the four tumours, although these two genes are
53
54 487 frequently mutated in luminal B tumours [54]. Despite the very limited sample size, we found that
55
56
57 488 *ATM*-associated tumours were more homogenous at the CNV level than at the SNV level. In
58
59 489 particular the two primary tumours from patient T0077 shared only 2.1% of SNVs. Moreover, no
60
61
62
63
64
65

490 specific mutation signature as defined by Alexandrov *et al.* (Nature 2013) could be evidenced using
1
2 491 SNV pattern of these four tumours only, and a larger tumour series should be sequenced to
3
4 492 determine whether such signatures can discriminate ATM-associated tumours.
5
6

7 493

9 494 **Conclusions**

10
11
12 495 Altogether, *ATM*-associated tumours do not show the HR deficiency profile and it is not clear
13
14 496 whether breast tumours developed by HetAT patients could be targeted by alkylating agents or PARP
15
16 497 inhibitors [55]. Nonetheless, hallmarks of *ATM*-associated tumours were found and could help to
17
18
19 498 identify *ATM* variant carriers outside of an A-T context or a HBOC family context. More studies are
20
21 499 needed to investigate whether genes located at loci 13q, 21p and 22q could harbour potential new
22
23
24 500 therapeutic targets and whether *RB1* deficiency could be a predictive biomarker for hormonotherapy
25
26 501 response for *ATM* patients.
27

28 502

30
31 503

33 504 **Abbreviations**

34
35
36 505 A-T: ataxia-telangiectasia; ATM: ataxia-telangiectasia mutated; BC: breast cancer; CNV: copy number
37
38 506 variation; ER: estrogen receptor; HetAT: ATM heterozygous carriers; HRD: homologous
39
40
41 507 recombination deficiency; HBOC: hereditary breast and ovarian cancers; LOH: loss of heterozygosity;
42
43 508 MS: missense substitution; Non-HetAT: noncarriers of an ATM variant; OR: odds ratio; PR:
44
45 509 progesterone receptor; SNV: single-nucleotide variant; TCGA: The Cancer Genome Atlas; TV:
46
47
48 510 truncating variant; WGS: whole-genome sequencing
49

50 511

52 53 512 **Declarations**

54 55 513 **Acknowledgements**

56
57
58 514 We are most grateful to the families who so willingly participated into the study. We acknowledge
59
60
61
62
63
64
65

515 David Gentien and Cécile Reyes for technical expertise. We thank all the CoF-AT collaborating cancer
1
2 516 clinics: Institut Curie, Paris: Dominique Stoppa-Lyonnet, Bruno Buecher, Antoine de Pauw and Sophie
3
4 517 Lejeune-Dumoulin; Hôpital Arnaud de Villeneuve, Montpellier: Isabelle Coupier; ICM Val d'Aurelle,
5
6
7 518 Montpellier: Isabelle Coupier; CHU de Nîmes: Audrey Combès; Centre François Baclesse, Caen:
8
9 519 Pascaline Berthet; Hôpital de la Timone, Marseille: Hélène Zattara; Clinique Universitaire Saint-Luc,
10
11 520 Bruxelles: Nicolas Janin and Karin Dahan; Centre Oscar Lambret, Lille: Philippe Vennin† and Claude
12
13
14 521 Adenis; Institut Bergonié, Bordeaux: Michel Longy; Centre Jean Perrin, Clermont-Ferrand: Jacques-
15
16 522 Olivier Bay; Centre Jean-Paul Strauss, Strasbourg: Jean-Pierre Fricker; Centre Catherine de Sienne,
17
18
19 523 Nantes: Alain Lortholary; CHU de Poitiers, Poitiers: Brigitte Gilbert-Dussardier; Hôpital d'Enfants,
20
21 524 Dijon: Laurence Faivre and Caroline Jacquot; Centre Antoine Lacassagne, Nice: Marc Frenay; Institut
22
23 525 Claudius Regaud, IUCT Oncopole, Toulouse: Laurence Gladieff; CHU de Grenoble, Grenoble:
24
25
26 526 Dominique Leroux; CHU de Lyon, Lyon: Gaétan Lesca; Gustave Roussy, Villejuif: Agnès Chompret†;
27
28 527 Centre Léon Bérard, Lyon: Christine Lasset; Hôtel-Dieu, Chambéry: James Lespinasse; CHI Toulon La
29
30
31 528 Seyne-sur-Mer, Toulon: Xavier Tchiknavorian; Centre Eugène Marquis, Rennes: Catherine Dugast†;
32
33 529 CHU du Mans, Le Mans: Dominique Martin-Coignard; CHRU Hôpital Caremeau, Nîmes: Jean Chiesa;
34
35 530 Hôpital Universitaire Dupuytren, Limoges: Laurence Venat-Bouvet; Centre René Gauducheau,
36
37
38 531 Nantes: Capucine Delnatte; Hôpital Porte-Madeleine, Orléans: Sonia Nizard; CHU de Nancy, Nancy:
39
40 532 Bruno Leheup; Hôpital Sainte-Musse, Toulon: Patrick Collignon; Polyclinique Courlancy, Reims: Liliane
41
42 533 Demange†; CHU de Besançon-Hôpital Jean Minjot, Besançon: Pierre Rohlrich; CHU Vaudois,
43
44
45 534 Lausanne: Florence Fellmann; CHU d'Angers, Angers: Isabelle Pellier; Hôpitaux de Rouen, Rouen:
46
47 535 Julie Tinat. We wish to thank more specifically our late colleague Josué Feingold† who has been
48
49 536 instrumental in the setup of the CoF-AT study. We are also thankful to the GENESIS study
50
51
52 537 collaborators and more specifically to Sylvie Mazoyer, Francesca Damiola and Laure Barjhoux who
53
54 538 managed the GENESIS biological resource until December 2015. We wish to thank Heather Thorne,
55
56
57 539 Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family
58
59 540 Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the
60
61
62
63
64
65

541 National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for
1
2 542 their contributions to this resource, and the many families who contribute to kConFab.
3

4
5 543 †Deceased prematurely.
6

7 544
8

9 10 545 **Fundings**

11
12
13 546 This work was supported by Inserm and Ministère de la Recherche (01P0751 - 01P0752 - 01P0753 -
14
15 547 01P0754 - 01P0755), Electricité de France (conseil scientifique de Radioprotection d'EDF, grant EP
16
17 548 2002-03, EP 2004-03, RB 2016-22), Fondation de France (grants 2001009761 and 2005011201), La
18
19
20 549 Ligue (grants PRE04/NA, PRE07/NA and PRE2015 LNCC/NA), La Ligue Comité du Maine et Loire,
21
22 550 MGEN Union, ITMO Santé Publique d'AVIESAN (grant AAP12-COH-110), Institut National du Cancer
23
24
25 551 (grant INCa-9578), the comprehensive cancer center SiRIC (Site de Recherche Intégrée sur le Cancer :
26
27 552 Grant INCa-DGOS-4654), Fondation ARC pour la recherche sur le cancer (grant PJA 20151203365)
28
29 553 and the Agence Nationale de la Recherche, program « investissements d'avenir » (grant ANR-10-
30
31
32 554 EQPX-03). ALR was the recipient of a PhD fellowship from the Fondation ARC pour la recherche sur le
33
34 555 cancer. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously
35
36 556 by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the
37
38
39 557 Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer
40
41 558 Foundation of Western Australia.
42

43
44 559
45

46 47 560 **Availability of data and materials**

48
49
50 561 The datasets generated during and/or analysed during the present study are not publicly
51
52
53 562 available, owing to confidentiality reasons, but anonymised versions may be available from
54
55 563 the corresponding author on reasonable request.
56

57
58 564
59
60
61
62
63
64
65

565 **Authors' contributions**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

566 Conception and design: ALR, JH, DSL, NA, FL.

567 Acquisition of data (invited and managed patients, provided facilities, etc.): EC, DLG, JB, SEM, ML, IB,
568 SB, JOB, PB, OC, BB, LF, MF, MG, V, PG, NJ, SL, CM, SM, LVB, HZ, JPF, LG, IC, GCT, DSL.

569 Analysis and interpretation of data (molecular analysis, statistical analysis, bioinformatics): ALR, NM,
570 LF, TP, EG, GB, CDE, JH, DSL, NA, FL.

571 Writing, review, and/or revision of the manuscript: ALR, LF, EG, JH, GCT, DSL, NA, FL.

572 Administrative, technical, or material support (i.e., reporting or organizing data, constructing
573 databases): ALR, NM, PLR, MGD, AL, WC, VR, ML, HT.

574 Study supervision: FL.

575
576 **Ethics approval and consent to participate**

577 Written informed consent was obtained from all participants. The Retro-AT, CoF-AT, and GENESIS
578 study protocols were approved by the appropriate ethics committee (CCP Ile-de-France III) and by
579 the French data protection authority (CNIL). The kConFab study was approved by Peter MacCallum
580 Cancer Centre Ethics Committee and Queensland Institute of Medical Research Human Research
581 Ethics Committee.

582
583 **Consent for publication**

584 Not applicable.

585
586 **Conflict of interest**

587 The authors declare no conflict of interest.

588

589 References

- 590 1. Micol R, Ben Slama L, Suarez F, Le Mignot L, Beaute J, Mahlaoui N, Dubois d'Enghien C, Lauge
591 A, Hall J, Couturier J *et al*: **Morbidity and mortality from ataxia-telangiectasia are associated**
592 **with ATM genotype.** *J Allergy Clin Immunol* 2011, **128**(2):382-389 e381.
- 593 2. Suarez F, Mahlaoui N, Canioni D, Andriamanga C, Dubois d'Enghien C, Brousse N, Jais JP,
594 Fischer A, Hermine O, Stoppa-Lyonnet D: **Incidence, presentation, and prognosis of**
595 **malignancies in ataxia-telangiectasia: a report from the French national registry of primary**
596 **immune deficiencies.** *J Clin Oncol* 2015, **33**(2):202-208.
- 597 3. Taylor AM, Metcalfe JA, Thick J, Mak YF: **Leukemia and lymphoma in ataxia telangiectasia.**
598 *Blood* 1996, **87**(2):423-438.
- 599 4. Borresen AL, Andersen TI, Tretli S, Heiberg A, Moller P: **Breast cancer and other cancers in**
600 **Norwegian families with ataxia-telangiectasia.** *Genes Chromosomes Cancer* 1990, **2**(4):339-
601 340.
- 602 5. Swift M: **Malignant disease in heterozygous carriers.** *Birth Defects Orig Artic Ser* 1976,
603 **12**(1):133-144.
- 604 6. Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, Byrd P, Taylor M, Easton DF:
605 **Cancer risks and mortality in heterozygous ATM mutation carriers.** *J Natl Cancer Inst* 2005,
606 **97**(11):813-822.
- 607 7. Andrieu N, Cavaciuti E, Lauge A, Ossian K, Janin N, Hall J, Stoppa-Lyonnet D: **Ataxia-**
608 **telangiectasia genes and breast cancer risk in a French family study.** *The Journal of dairy*
609 *research* 2005, **72 Spec No**:73-80.
- 610 8. Renault AL, Mebirouk N, Cavaciuti E, Le Gal D, Lecarpentier J, d'Enghien CD, Lauge A, Dondon
611 MG, Labbe M, Lesca G *et al*: **Telomere length, ATM mutation status and cancer risk in**
612 **Ataxia-Telangiectasia families.** *Carcinogenesis* 2017.
- 613 9. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot
614 R, Spanova K *et al*: **ATM mutations that cause ataxia-telangiectasia are breast cancer**
615 **susceptibility alleles.** *Nat Genet* 2006, **38**(8):873-875.
- 616 10. Tavgigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F,
617 Byrnes GB, Chuang SC, Forey N *et al*: **Rare, evolutionarily unlikely missense substitutions in**
618 **ATM confer increased risk of breast cancer.** *Am J Hum Genet* 2009, **85**(4):427-446.
- 619 11. Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavgigian SV, Nathanson KL, Devilee P,
620 Meindl A, Couch FJ, Southey M *et al*: **Gene-panel sequencing and the prediction of breast-**
621 **cancer risk.** *N Engl J Med* 2015, **372**(23):2243-2257.

- 622 12. Balleine RL, Murali R, Bilous AM, Farshid G, Waring P, Provan P, Byth K, Thorne H, kConFab I,
1 623 Kirk JA: **Histopathological features of breast cancer in carriers of ATM gene variants.**
2 624 *Histopathology* 2006, **49**(5):523-532.
3
- 4 625 13. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, Terry MB, Daly MJ, Buys SM,
5 626 Southey MC *et al*: **Rare variants in the ATM gene and risk of breast cancer.** *Breast cancer*
6 627 *research : BCR* 2011, **13**(4):R73.
8
- 9 628 14. Waddell N, Cocciardi S, Johnson J, Healey S, Marsh A, Riley J, da Silva L, Vargas AC, Reid L,
10 629 kConFab I *et al*: **Gene expression profiling of formalin-fixed, paraffin-embedded familial**
11 630 **breast tumours using the whole genome-DASL assay.** *J Pathol* 2010, **221**(4):452-461.
12
- 13 631 15. Bubien V, Bonnet F, Dupiot-Chiron J, Barouk-Simonet E, Jones N, de Reynies A, MacGrogan G,
14 632 Sevenet N, Letouze E, Longy M: **Combined tumor genomic profiling and exome sequencing**
15 633 **in a breast cancer family implicates ATM in tumorigenesis: a proof of principle study.** *Genes*
16 634 *Chromosomes Cancer* 2017.
18
- 19 635 16. Popova T, Manie E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-
20 636 Zafrani B, Bollet M, Longy M *et al*: **Ploidy and large-scale genomic instability consistently**
21 637 **identify basal-like breast carcinomas with BRCA1/2 inactivation.** *Cancer Res* 2012,
22 638 **72**(21):5454-5462.
24
- 25 639 17. Janin N, Andrieu N, Ossian K, Lauge A, Croquette MF, Griscelli C, Debre M, Bressac-de-
26 640 Paillerets B, Aurias A, Stoppa-Lyonnet D: **Breast cancer risk in ataxia telangiectasia (AT)**
27 641 **heterozygotes: haplotype study in French AT families.** *Br J Cancer* 1999, **80**(7):1042-1045.
29
- 30 642 18. Sinilnikova OM, Dondon MG, Eon-Marchais S, Damiola F, Barjhoux L, Marcou M, Verny-Pierre
31 643 C, Sornin V, Toulemonde L, Beauvallet J *et al*: **GENESIS: a French national resource to study**
32 644 **the missing heritability of breast cancer.** *BMC Cancer* 2016, **16**:13.
33
- 34 645 19. Mann GJ, Thorne H, Balleine RL, Butow PN, Clarke CL, Edkins E, Evans GM, Fereday S, Haan E,
35 646 Gattas M *et al*: **Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the**
36 647 **kConFab familial breast cancer resource.** *Breast cancer research : BCR* 2006, **8**(1):R12.
38
- 39 648 20. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB, Group IUGVW: **In silico analysis of**
40 649 **missense substitutions using sequence-alignment based methods.** *Hum Mutat* 2008,
41 650 **29**(11):1327-1336.
42
- 43 651 21. Elston CW, Ellis IO, Pinder SE: **Pathological prognostic factors in breast cancer.** *Critical*
44 652 *reviews in oncology/hematology* 1999, **31**(3):209-223.
46
- 47 653 22. Bocker W: **[WHO classification of breast tumors and tumors of the female genital organs:**
48 654 **pathology and genetics].** *Verh Dtsch Ges Pathol* 2002, **86**:116-119.
49
- 50 655 23. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ,
51 656 Panel m: **Personalizing the treatment of women with early breast cancer: highlights of the**
52 657 **St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer**
53 658 **2013.** *Ann Oncol* 2013, **24**(9):2206-2223.
55
- 56 659 24. Popova T, Manie E, Stoppa-Lyonnet D, Rigaiil G, Barillot E, Stern MH: **Genome Alteration**
57 660 **Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP**
58 661 **arrays.** *Genome biology* 2009, **10**(11):R128.
60
61
62
63
64
65

- 662 25. La Rosa P, Viara E, Hupe P, Pierron G, Liva S, Neuvial P, Brito I, Lair S, Servant N, Robine N *et al*: **VAMP: visualization and analysis of array-CGH, transcriptome and other molecular profiles**. *Bioinformatics* 2006, **22**(17):2066-2073.
- 1 663
- 2 664
- 3
- 4 665 26. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
- 5 666
- 6
- 7
- 8 667 27. Shen R, Seshan VE: **FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing**. *Nucleic Acids Res* 2016, **44**(16):e131.
- 9 668
- 10
- 11 669 28. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: **VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing**. *Genome Res* 2012, **22**(3):568-576.
- 12 670
- 13 671
- 14
- 15
- 16 672 29. Cancer Genome Atlas N: **Comprehensive molecular portraits of human breast tumours**. *Nature* 2012, **490**(7418):61-70.
- 17 673
- 18
- 19 674 30. Valla M, Vatten LJ, Engstrom MJ, Haugen OA, Akslen LA, Bjorngaard JH, Hagen AI, Ytterhus B, Bofin AM, Opdahl S: **Molecular Subtypes of Breast Cancer: Long-term Incidence Trends and Prognostic Differences**. *Cancer Epidemiol Biomarkers Prev* 2016, **25**(12):1625-1634.
- 20 675
- 21 676
- 22
- 23
- 24 677 31. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J *et al*: **A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer**. *Clin Cancer Res* 2010, **16**(21):5222-5232.
- 25 678
- 26 679
- 27 680
- 28
- 29
- 30 681 32. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J *et al*: **Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer**. *J Clin Oncol* 2006, **24**(23):3726-3734.
- 31 682
- 32 683
- 33
- 34 684 33. Manie E, Popova T, Battistella A, Tarabeux J, Caux-Moncoutier V, Golmard L, Smith NK, Mueller CR, Mariani O, Sigal-Zafrani B *et al*: **Genomic hallmarks of homologous recombination deficiency in invasive breast carcinomas**. *Int J Cancer* 2016, **138**(4):891-900.
- 35 685
- 36 686
- 37
- 38
- 39 687 34. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups**. *Nature* 2012, **486**(7403):346-352.
- 40 688
- 41 689
- 42
- 43 690 35. Pecuchet N, Popova T, Manie E, Lucchesi C, Battistella A, Vincent-Salomon A, Caux-Moncoutier V, Bollet M, Sigal-Zafrani B, Sastre-Garau X *et al*: **Loss of heterozygosity at 13q13 and 14q32 predicts BRCA2 inactivation in luminal breast carcinomas**. *Int J Cancer* 2013, **133**(12):2834-2842.
- 44 691
- 45 692
- 46 693
- 47
- 48
- 49 694 36. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. *Cancer Discov* 2012, **2**(5):401-404.
- 50 695
- 51 696
- 52
- 53
- 54 697 37. Decker B, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK, Wang Q, Ahmed S, Baynes C, Conroy DM *et al*: **Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks**. *J Med Genet* 2017, **54**(11):732-741.
- 55 698
- 56 699
- 57 700
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 701 38. Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, Szwiec M,
1 702 Byrski T, Debniak T, Gorski B *et al*: **Clinical outcomes in women with breast cancer and a**
2 703 **PALB2 mutation: a prospective cohort analysis.** *Lancet Oncol* 2015, **16**(6):638-644.
3
4 704 39. Musolino A, Bella MA, Bortesi B, Michiara M, Naldi N, Zanelli P, Capelletti M, Pezzuolo D,
5 705 Camisa R, Savi M *et al*: **BRCA mutations, molecular markers, and clinical variables in early-**
6 706 **onset breast cancer: a population-based study.** *Breast* 2007, **16**(3):280-292.
8
9 707 40. Cybulski C, Huzarski T, Byrski T, Gronwald J, Debniak T, Jakubowska A, Gorski B, Wokolorczyk
10 708 D, Masojc B, Narod SA *et al*: **Estrogen receptor status in CHEK2-positive breast cancers:**
11 709 **implications for chemoprevention.** *Clin Genet* 2009, **75**(1):72-78.
12
13 710 41. Keeney MG, Couch FJ, Visscher DW, Lindor NM: **Non-BRCA familial breast cancer: review of**
14 711 **reported pathology and molecular findings.** *Pathology* 2017, **49**(4):363-370.
16
17 712 42. Massink MP, Kooi IE, Martens JW, Waisfisz Q, Meijers-Heijboer H: **Genomic profiling of**
18 713 **CHEK2*1100delC-mutated breast carcinomas.** *BMC Cancer* 2015, **15**:877.
19
20 714 43. Mavaddat N, Barrowdale D, Andrulis IL, Domchek SM, Eccles D, Nevanlinna H, Ramus SJ,
21 715 Spurdle A, Robson M, Sherman M *et al*: **Pathology of breast and ovarian cancers among**
22 716 **BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of**
23 717 **Modifiers of BRCA1/2 (CIMBA).** *Cancer Epidemiol Biomarkers Prev* 2012, **21**(1):134-147.
25
26 718 44. Kaurah P, MacMillan A, Boyd N, Senz J, De Luca A, Chun N, Suriano G, Zaor S, Van Manen L,
27 719 Gilpin C *et al*: **Founder and recurrent CDH1 mutations in families with hereditary diffuse**
28 720 **gastric cancer.** *JAMA* 2007, **297**(21):2360-2372.
30
31 721 45. Banneau G, Guedj M, MacGrogan G, de Mascarel I, Velasco V, Schiappa R, Bonadona V, David
32 722 A, Dugast C, Gilbert-Dussardier B *et al*: **Molecular apocrine differentiation is a common**
33 723 **feature of breast cancer in patients with germline PTEN mutations.** *Breast cancer research :*
34 724 *BCR* 2010, **12**(4):R63.
35
36 725 46. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K,
37 726 Broaddus R, Lu KH, Chen J *et al*: **Patterns of genomic loss of heterozygosity predict**
38 727 **homologous recombination repair defects in epithelial ovarian cancer.** *Br J Cancer* 2012,
39 728 **107**(10):1776-1782.
41
42 729 47. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, Rosebrock D, Livitz D, Kubler K,
43 730 Mouw KW *et al*: **A mutational signature reveals alterations underlying deficient**
44 731 **homologous recombination repair in breast cancer.** *Nat Genet* 2017, **49**(10):1476-1486.
46
47 732 48. Sagona AP, Stenmark H: **Cytokinesis and cancer.** *FEBS Lett* 2010, **584**(12):2652-2661.
48
49 733 49. Jonsdottir AB, Stefansson OA, Bjornsson J, Jonasson JG, Ogmundsdottir HM, Eyfjord JE:
50 734 **Tetraploidy in BRCA2 breast tumours.** *Eur J Cancer* 2012, **48**(3):305-310.
52
53 735 50. di Iasio MG, Calin G, Tibiletti MG, Vorechovsky I, Benediktsson KP, Taramelli R, Barbanti-
54 736 Brodano G, Negrini M: **Refinement of the LOH region 1 at 11q23.1 deleted in human breast**
55 737 **carcinomas and sublocalization of 11 expressed sequence tags within the refined region.**
56 738 *Oncogene* 1999, **18**(8):1635-1638.
57
58
59
60
61
62
63
64
65

- 739 51. Rio PG, Pernin D, Bay JO, Albuissou E, Kwiatkowski F, De Latour M, Bernard-Gallon DJ, Bignon
 1 740 YJ: **Loss of heterozygosity of BRCA1, BRCA2 and ATM genes in sporadic invasive ductal**
 2 741 **breast carcinoma.** *Int J Oncol* 1998, **13**(4):849-853.
- 4 742 52. Knudsen ES, Knudsen KE: **Tailoring to RB: tumour suppressor status and therapeutic**
 5 743 **response.** *Nat Rev Cancer* 2008, **8**(9):714-724.
- 8 744 53. Cho JS, Park MH, Lee JS, Yoon JH: **Reduced MUC4 expression is a late event in breast**
 9 745 **carcinogenesis and is correlated with increased infiltration of immune cells as well as**
 10 746 **promoter hypermethylation in invasive breast carcinoma.** *Appl Immunohistochem Mol*
 11 747 *Morphol* 2015, **23**(1):44-53.
- 14 748 54. Marchio C, Geyer FC, Ng CK, Piscuoglio S, De Filippo MR, Cupo M, Schultheis AM, Lim RS,
 15 749 Burke KA, Guerini-Rocco E *et al*: **The genetic landscape of breast carcinomas with**
 16 750 **neuroendocrine differentiation.** *J Pathol* 2017, **241**(3):405-419.
- 18 751 55. Choi M, Kipps T, Kurzrock R: **ATM Mutations in Cancer: Therapeutic Implications.** *Mol*
 19 752 *Cancer Ther* 2016, **15**(8):1781-1791.

20 753
 21
 22
 23 754
 24
 25 755
 26
 27

28 756 **Figure legends**

30 757 **Figure 1. Distribution of molecular subtypes in the ATM series and in the 3 control series.** PAM50
 31
 32
 33 758 classification was used for TCGA data, which explains the absence of luminal B/HER2+ tumours in this
 34
 35 759 series. A Fisher exact test was used to assess difference between ATM series and control series.

37 760
 38
 39
 40 761 **Figure 2. Copy number variation profiles of ATM-associated tumours analysed with the OncoScan**
 41
 42 762 **array. A.** Genome-wide view of cumulative CNVs present in the 23 ATM-associated tumours. Gains
 43
 44
 45 763 are indicated in red, losses in blue, and LOH in orange. **B.** Cluster dendrogram and genomic regions
 46
 47 764 altered in at least 70% of the 23 analysed tumours. AT children are indicated by a star (*). **C.** Cluster
 48
 49 765 dendrogram and genomic regions altered in at least 70% of the 16 tumours with confirmed biallelic
 50
 51
 52 766 inactivation of ATM. AT children are indicated by a star (*). Loss: the two alleles are present in the
 53
 54 767 tumour; Loss/LOH: only one allele is present in the tumour; Loss/LOH or Loss: consecutive segmental
 55
 56 768 regions characterized as either “Loss/LOH” or “Loss”.

58
 59 769
 60
 61
 62
 63
 64
 65

770 **Figure 3. CNV and SNV profiles of *ATM*-associated tumours obtained by deep whole-genome**
1
2 771 **sequencing. A.** Cumulative profiles of copy number gains, losses and of LOH regions obtained from
3
4 772 whole-genome sequencing (WGS) of 4 *ATM*-associated tumours. Dark boxes indicate the genomic
5
6
7 773 regions identified in the OncoScan analysis; green boxes indicated the new genomic regions
8
9 774 identified by WGS. **B.** Venn diagram representing the number of somatic SNVs and indels shared
10
11 775 between the 4 tumours. **C.** Venn diagram representing the number of genes altered and shared
12
13
14 776 between the 4 tumours.

15
16 777

17
18
19 778 **Tables list**

20
21 779 **Table 1.** Clinical characteristics of *ATM* mutation carriers, and available tumour material used for
22
23
24 780 analyses.

25
26 781 **Table 2.** Clinical and histological features of *ATM*-associated invasive breast carcinomas compared
27
28 782 with those of sporadic cases.

29
30
31 783 **Table 3.** Copy number losses recurrently observed in the 23 *ATM*-associated breast tumours.

32
33 784 **Supplementary Table S1.** Genomic regions showing copy number loss or loss/LOH in at least 70% of
34
35 785 *ATM*-associated tumours.

36
37
38 786 **Supplementary Table S2.** Comparison of genes altered at the CN level between *ATM*-associated
39
40 787 tumours with tumours from TCGA.

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Table 1.** Clinical characteristics of *ATM* mutation carriers, and available tumour material used for analyses.

2

Study	Patient ID	Sex	Nucleotide change	Effect on protein	Variant type*	Tumour ID	Age at diagnosis	Stade	Neoadjuvant treatment	Conservation	OncoScan Analysis
A-T families	AT1	M	c.2839-580_577del4(-/-)	cryptic splice site	TV ⁽¹⁾	T0072-L	28	Invasive	Unknown	FFPE	Yes
(Retro-AT + CoF-AT)	AT2	F	c.8585-2A>C	frameshift	TV	T0075-L	42	Invasive	Unknown	Bouin	No
			c.5189G>T	p.Arg1730Leu	MS (C0)						
	AT3	F	c.2413C>T	p.Arg805X	TV	T0249-L	30	Invasive	Unknown	FFPE	Yes
			c.7517_7520delGAGA	p.Arg2506ThrfsX3							
	1	F	c.3576G>A	p.Ser1135_Lys1192 del58	TV	T0001-L	62	Invasive	No	FFPE+Frozen	No
	2	F	c.2839-580_577del4	cryptic splice site	TV	T0002-L	60	Invasive	No	Bouin	No
	3	F	c.5644C>T	p.Arg1882X	TV	T0003-L	45	Invasive	No	FFPE	No
	4	F	c.3802delG	p.Val1268X	TV	T0005-L	65	Invasive	No	FFPE	Yes
	5	F	c.6007-2A>T	frameshift	TV	T0007-R	74	Invasive	No	FFPE	No
	6	F	c.3085dupA	p.Thr1029AsnfsX19	TV	T0008-R	51	Invasive	Unknown	FFPE	No
	7	F	c.3894dupT	p.Ala1299CysfsX3	TV	T0009-L	36	Invasive	No	FFPE	Yes
	8	F	c.6007-2A>T	frameshift	TV	T0010-L	30	Invasive	Unknown	Bouin	No
	9	F	c.6404_6405insTT	p.Arg2136X	TV	T0015-L	40	Invasive	No	FFPE+Frozen	Yes
	10	F	c.2466fs	del exon19-65	TV	T0016-R	62	Invasive	No	Bouin	No

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

						T0016-L	72	Invasive	No	FFPE	Yes
	11	F	c.3153+1G>A	frameshift	TV	T0073-L	77	Invasive	Unknown	Bouin	No
	12	F	c.8489T>G	p.Val2830Gly	MS (C65)	T0074-L	62	Invasive	Unknown	Bouin	No
	13	F	c.73-2A>G	frameshift	TV	T0076-R	47	Invasive	No	FFPE	Yes
	14	F	c.3754_3756delTATinsCA	p.Met2918IlefsX21	TV	T0077-L	66	Invasive	No	FFPE+Frozen	Yes
						T0077-R	66	Invasive	No	FFPE+Frozen	Yes
	15	F	c.8140C>T	p.Gln2714X	TV	T0078-L	55	Invasive	No	FFPE	Yes
	16	F	c.5644C>T	p.Arg1882X	TV	T0181-R	39	In situ	Unknown	FFPE	No
	17	F	c.8083G>A	p.Gly2695Ser	MS (C55)	T0247-R	48	Invasive	No	FFPE	Yes
	18	F	c.7928-2A>C	frameshift	TV	T0248-L	35	Invasive	Unknown	FFPE	Yes
GENESIS	19	F	c.2413C>T	p.Arg805X	TV	T0045-R	31	Invasive	No	No material	-
						T0045-R	67	In situ	No	FFPE	Yes
	20	F	c.8584+1G>A	frameshift	TV	T0091-L	51	In situ	No	FFPE	No
	21	F	c.3058dupA	p.Leu1019fs	TV	T0099-L	39	Invasive	No	FFPE	Yes
	22	F	c.5497-2A>C	p.Val1833IlefsX	TV ⁽²⁾	T0111-L	32	Invasive	No	FFPE	No
	23	F	c.9008A>T	p.Asn3003Ile	MS (C65)	T0118-L	64	Invasive	No	FFPE	No
	24	F	c.1464G>T	p.Trp488Cys	MS (C65)	T0120-R	65	Invasive	No	FFPE	Yes
	25	F	c.5527delC	p.Phe1843fs	TV	T0123-R	74	Invasive	No	FFPE	Yes
	26	F	c.5750G>C	p.Arg1917Thr	MS (C65)	T0191-L	45	Invasive	Unknown	FFPE	No
	27	F	c.1236-2A>T	p.Trp412X	TV	T0192-R	42	In situ	Unknown	FFPE	No

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

	28	F	c.8614C>A	p.His2872Asn	MS (C65)	T0218-R	54	Invasive	Unknown	FFPE	No
	29	F	c.8494C>T	p.Arg2832Cys	MS ⁽³⁾ (C45)	T0220-R	42	Invasive	No	FFPE	Yes
kConFab	30	F	c.3801delG	p.Glu1267fs	TV	T0173-L	44	Invasive	Unknown	No material	-
						T0173-R	49	Invasive	Unknown	FFPE	Yes
	31	M	c.6820G>A	p.Ala2274Thr	MS (C55)	T0174-R	45	In situ	Unknown	FFPE	Yes
	32	F	c.4909+1G>A	frameshift	TV ⁽⁴⁾	T0175-R	42	Invasive	Unknown	FFPE	Yes
	33	F	c.8266A>T	p.Lys2756X	TV ⁽⁵⁾	T0176-R	60	Invasive	Unknown	FFPE	Yes
	34	F	c.8158G>C	p.Asp2720His	MS (C65)	T0177-L	59	Invasive	Unknown	FFPE	No
	35	F	c.8266A>T	p.Lys2756X	TV ⁽⁵⁾	T0179-R	41	Invasive	Unknown	No material	-
						T0179-L	50	Invasive	Unknown	FFPE	Yes
	36	F	c.7176_7177insT	p.Ser2394PhefsX9	TV	T0180-R	43	Invasive	Unknown	FFPE	Yes

1
2
3
4
5

AT, ataxia-telangiectasia; F, female; M, male; TV, truncating variant; MS, missense substitution. L, left, R, right; FFPE, formalin-fixed, paraffin-embedded tissue sample; Bouin, Bouin-fixed, paraffin-embedded tissue sample.
(1,2,3,4,5) Reported as pathogenic for A-T in ClinVar.
* Align-GVGD grades are indicated in brackets for MS variants

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Table 2.** Clinical and histological features of *ATM*-associated invasive breast carcinomas compared with those of sporadic cases.

Clinicopathological variable	PICBIM series (N=516)	ATM series (all tumours, TV+MS) (N=36)	<i>P</i> ^a	ATM series (1st primary tumours only) (N=32)	<i>P</i> ^a	ATM series (excluding MS in HBOC families) (N=31)	<i>P</i> ^a
Histological subtype							
Ductal carcinoma	434	30	Ref.	27	Ref.	25	Ref.
Lobular carcinoma	54	2	0.55	2	0.69	2	0.78
Others	28	3	0.43	2	0.78	3	0.25
Unknown	0	1	-	1	-	1	-
Architecture							
1-2	126	0	Ref.	7	Ref.	6	Ref.
3	390	30	0.97	21	0.86	20	0.96
Unknown	0	6	-	4	-	5	-
Mitosis							

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

0-1	221	15	Ref.	14	Ref.	12	Ref.
2	103	9	0.58	8	0.70	9	0.31
3	191	6	0.10	6	0.13	5	0.15
Unknown	1	6	-	4	-	5	-
Anisokaryosis							
1	30	2	Ref.	2	Ref.	2	Ref.
2	225	7	0.25	7	0.22	5	0.12
3	261	21	0.96	19	0.89	19	0.9
Unknown	0	6	-	4	-	5	-
Histological grade							
I	88	4	Ref.	3	Ref.	4	Ref.
II	191	16	0.34	15	0.24	13	0.58
III	236	15	0.64	13	0.55	14	0.75
Unknown	1	1	-	1	-	-	-

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Tumour size (cm)

pT1 (<2)	329	23	Ref.	20	Ref.	19	Ref.
pT2 (2 - 5)	166	8	0.37	7	0.41	7	0.48
pT3 (>5)	14	2	0.37	2	0.30	2	0.26
pT4	7	0	-	0	-	0	-
Unknown	0	3	-	3	-	3	-

Pushing margins

Absent	461	19	Ref.	17	Ref.	15	Ref.
Present	49	2	0.98	2	0.93	2	0.78
Unknown	6	15	-	13	-	14	-

Emboles

Absent	315	13	Ref.	11	Ref.	13	Ref.
Present	199	14	0.08	14	0.04	11	0.32

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Unknown	2	9	-	7	-	7	-
N stage							
pN0	286	18	Ref.	15	Ref.	15	Ref.
pN1	151	10	0.80	9	0.66	9	0.55
pN2	59	1	0.23	1	0.32	1	0.35
pN3	16	1	0.96	1	0.81	1	0.76
pNx	5	6	-	6	-	5	-
Estrogen receptor							
Positive	307	34	Ref.	31	Ref.	29	Ref.
Negative	209	1	0.003	1	0.004	1	0.005
Unknown	0	1	-	-	-	1	-
Progesterone receptor							
Positive	278	26	Ref.	23	Ref.	23	Ref.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Negative	236	8	0.03	8	0.07	6	0.02
Unknown	2	2	-	1	-	2	-
HER2							
Negative	433	25	Ref.	23	Ref.	21	Ref.
Positive	83	6	0.77	5	0.97	6	0.53
Unknown	0	5	-	4	-	4	-
Ki67							
<20%	173	11	Ref.	10	Ref.	8	Ref.
≥20%	336	18	0.55	17	0.59	17	0.97
Unknown	7	7	-	5	-	6	-
Molecular subtype							
TNBC	142	1	Ref.	1	Ref.	1	Ref.
HER2	66	0	N/A	0	N/A	0	N/A

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Luminal A	180	10	0.06	9	0.08	7	0.13
Luminal B	111	13	0.009	12	0.010	12	0.010
Luminal B / HER2	17	4	0.005	4	0.005	4	0.005
Unknown	0	8	-	6	-	7	-

1 ^a P-value adjusted for gender and age at diagnosis

2 TNBC, triple-negative breast cancer

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Table 3.** Copy number losses recurrently observed in the 23 *ATM*-associated breast tumours.

2

Locus	Associated morphology	All ATM tumours (N=23)			ATM tumours with proven biallelic inactivation of ATM (N=16)			ATM tumours sequenced by WGS (N=4)		
		Cytogenetic band	Number of genes	Cancer genes*	Cytogenetic band	Number of genes	Cancer genes*	Cytogenetic band	Number of genes	Cancer genes*
6q	20% of ATM tumours							6q23.3-q27	190	<i>ARID1B, ECT2L, ESR1, EZR, FGFR10P, QKI, TNFAIP3</i>
8p	Breast cancer	8p21.3	6 ^d	-	8p23.3-p12	216	<i>WRN, NRG1</i>	8p21.3	6 ^d	-
11q	ATM tumours	11q22	2	<i>ATM, ZBTB16</i>	11q21-q25	276	<i>ATM, ZBTB16, DDX10, POU2AF1, SDHD, ZBTB16, PAFAH1B2,</i>	11q21-q24.2	236	<i>ATM, ZBTB16, DDX10, POU2AF1, SDHD, ZBTB16, PAFAH1B2,</i>

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

							<i>PDCK7, KMT2A,</i>			<i>PDCK7, KMT2A,</i>
							<i>MAML2, DDX6,</i>			<i>MAML2, DDX6,</i>
							<i>BCL9L, CBL,</i>			<i>BCL9L, CBL</i>
							<i>ARHGEF12, KCNJ5</i>			
13q	70% ATM tumours	13q14.11- q14.3	90	<i>LCP1, RB1</i>	13q13.3- q32.3	320	<i>LHFP, FOXO1,</i> <i>LCP1, RB1</i>	13q13.3- q32.3	320	<i>LHFP, FOXO1,</i> <i>LCP1, RB1</i>
16q	Luminal tumours	16q13-q24.3	364	<i>HERPUD1,</i> <i>CDH11, CBF,</i> <i>CTCF, CDH1,</i> <i>ZFH3, MAF,</i> <i>CBFA2T3,</i> <i>FANCA</i>	16q22.1	1 ^c	-	16q22.1	1 ^c	-
17p	Luminal tumours	17p13.3	46	<i>YWHAE</i>		-	-	17p13.3	46	<i>YWHAE</i>
17p	Luminal tumours	17p13.2-p12	166	<i>USP6,</i> <i>RABEP1,</i>		-	-			

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

TP53, GAS7,
MAP2K4

19p	40% of ATM tumours							19p13.3- p13.2	256	<i>FSTL3, GNA11,</i> <i>MAP2K2, MLLT1,</i> <i>SH3GL1, STK11,</i> <i>TCF3</i>
-----	-----------------------	--	--	--	--	--	--	-------------------	-----	---

21p	70% of ATM tumours	21p11.2- p11.1	2 ^a	-	21p11.2- p11.1	2 ^a	-			
-----	-----------------------	-------------------	----------------	---	-------------------	----------------	---	--	--	--

22q	Luminal B tumours, 70% ATM tumours	22q11.23	2 ^b		22q11.23	3 ^b		<i>GSTT1</i>		
	Luminal B tumours, 70% ATM tumours	-	-	-	22q12.3	199		<i>APOBEC3B, MKL1,</i> <i>EP300</i>		

1 * as reported in the COSMIC database.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 ^aThis region contains TEKT4P2, TPTE and 4 miRNAs (MIR3648-1, MIR3648-2, MIR3687-1, MIR3687-2).

2 ^bThis region contains the pseudogenes *GSTTP1* and *GSTTP2*.

3 ^cThis region contains WWP2 and the miRNA MIR140.

4 ^dThis region contains *PPP3CC*, *SORBS3*, *PDLIM2*, *BIN3*, *BIN3-IT1*, *EGR3*.

5

1 **Table 1.** Clinical characteristics of *ATM* mutation carriers, and available tumour material used for analyses.

2

Study	Patient ID	Sex	Nucleotide change	Effect on protein	Variant type*	Tumour ID	Age at diagnosis	Stade	Neoadjuvant treatment	Conservation	OncoScan Analysis
A-T families	AT1	M	c.2839-580_577del4(-/-)	cryptic splice site	TV ⁽¹⁾	T0072-L	28	Invasive	Unknown	FFPE	Yes
(Retro-AT + CoF-AT)	AT2	F	c.8585-2A>C	frameshift	TV	T0075-L	42	Invasive	Unknown	Bouin	No
			c.5189G>T	p.Arg1730Leu	MS (C0)						
	AT3	F	c.2413C>T	p.Arg805X	TV	T0249-L	30	Invasive	Unknown	FFPE	Yes
			c.7517_7520delGAGA	p.Arg2506ThrfsX3							
	1	F	c.3576G>A	p.Ser1135_Lys1192 del58	TV	T0001-L	62	Invasive	No	FFPE+Frozen	No
	2	F	c.2839-580_577del4	cryptic splice site	TV	T0002-L	60	Invasive	No	Bouin	No
	3	F	c.5644C>T	p.Arg1882X	TV	T0003-L	45	Invasive	No	FFPE	No
	4	F	c.3802delG	p.Val1268X	TV	T0005-L	65	Invasive	No	FFPE	Yes
	5	F	c.6007-2A>T	frameshift	TV	T0007-R	74	Invasive	No	FFPE	No
	6	F	c.3085dupA	p.Thr1029AsnfsX19	TV	T0008-R	51	Invasive	Unknown	FFPE	No
	7	F	c.3894dupT	p.Ala1299CysfsX3	TV	T0009-L	36	Invasive	No	FFPE	Yes
	8	F	c.6007-2A>T	frameshift	TV	T0010-L	30	Invasive	Unknown	Bouin	No
	9	F	c.6404_6405insTT	p.Arg2136X	TV	T0015-L	40	Invasive	No	FFPE+Frozen	Yes
	10	F	c.2466fs	del exon19-65	TV	T0016-R	62	Invasive	No	Bouin	No

						T0016-L	72	Invasive	No	FFPE	Yes
	11	F	c.3153+1G>A	frameshift	TV	T0073-L	77	Invasive	Unknown	Bouin	No
	12	F	c.8489T>G	p.Val2830Gly	MS (C65)	T0074-L	62	Invasive	Unknown	Bouin	No
	13	F	c.73-2A>G	frameshift	TV	T0076-R	47	Invasive	No	FFPE	Yes
	14	F	c.3754_3756delTATinsCA	p.Met2918IlefsX21	TV	T0077-L	66	Invasive	No	FFPE+Frozen	Yes
						T0077-R	66	Invasive	No	FFPE+Frozen	Yes
	15	F	c.8140C>T	p.Gln2714X	TV	T0078-L	55	Invasive	No	FFPE	Yes
	16	F	c.5644C>T	p.Arg1882X	TV	T0181-R	39	In situ	Unknown	FFPE	No
	17	F	c.8083G>A	p.Gly2695Ser	MS (C55)	T0247-R	48	Invasive	No	FFPE	Yes
	18	F	c.7928-2A>C	frameshift	TV	T0248-L	35	Invasive	Unknown	FFPE	Yes
GENESIS	19	F	c.2413C>T	p.Arg805X	TV	T0045-R	31	Invasive	No	No material	-
						T0045-R	67	In situ	No	FFPE	Yes
	20	F	c.8584+1G>A	frameshift	TV	T0091-L	51	In situ	No	FFPE	No
	21	F	c.3058dupA	p.Leu1019fs	TV	T0099-L	39	Invasive	No	FFPE	Yes
	22	F	c.5497-2A>C	p.Val1833IlefsX	TV ⁽²⁾	T0111-L	32	Invasive	No	FFPE	No
	23	F	c.9008A>T	p.Asn3003Ile	MS (C65)	T0118-L	64	Invasive	No	FFPE	No
	24	F	c.1464G>T	p.Trp488Cys	MS (C65)	T0120-R	65	Invasive	No	FFPE	Yes
	25	F	c.5527delC	p.Phe1843fs	TV	T0123-R	74	Invasive	No	FFPE	Yes
	26	F	c.5750G>C	p.Arg1917Thr	MS (C65)	T0191-L	45	Invasive	Unknown	FFPE	No
	27	F	c.1236-2A>T	p.Trp412X	TV	T0192-R	42	In situ	Unknown	FFPE	No

	28	F	c.8614C>A	p.His2872Asn	MS (C65)	T0218-R	54	Invasive	Unknown	FFPE	No
	29	F	c.8494C>T	p.Arg2832Cys	MS ⁽³⁾	T0220-R	42	Invasive	No	FFPE	Yes
					(C45)						
kConFab	30	F	c.3801delG	p.Glu1267fs	TV	T0173-L	44	Invasive	Unknown	No material	-
						T0173-R	49	Invasive	Unknown	FFPE	Yes
	31	M	c.6820G>A	p.Ala2274Thr	MS (C55)	T0174-R	45	In situ	Unknown	FFPE	Yes
	32	F	c.4909+1G>A	frameshift	TV ⁽⁴⁾	T0175-R	42	Invasive	Unknown	FFPE	Yes
	33	F	c.8266A>T	p.Lys2756X	TV ⁽⁵⁾	T0176-R	60	Invasive	Unknown	FFPE	Yes
	34	F	c.8158G>C	p.Asp2720His	MS (C65)	T0177-L	59	Invasive	Unknown	FFPE	No
	35	F	c.8266A>T	p.Lys2756X	TV ⁽⁵⁾	T0179-R	41	Invasive	Unknown	No material	-
						T0179-L	50	Invasive	Unknown	FFPE	Yes
	36	F	c.7176_7177insT	p.Ser2394PhefsX9	TV	T0180-R	43	Invasive	Unknown	FFPE	Yes

1

2 AT, ataxia-telangiectasia; F, female; M, male; TV, truncating variant; MS, missense substitution. L, left, R, right; FFPE, formalin-fixed, paraffin-embedded

3 tissue sample; Bouin, Bouin-fixed, paraffin-embedded tissue sample.

4 (1,2,3,4,5) Reported as pathogenic for A-T in ClinVar.

5 * Align-GVGD grades are indicated in brackets for MS variants

1 **Table 2.** Clinical and histological features of *ATM*-associated invasive breast carcinomas compared with those of sporadic cases.

Clinicopathological variable	PICBIM series (N=516)	ATM series (all tumours, TV+MS) (N=36)	<i>P</i> ^a	ATM series (1st primary tumours only) (N=32)	<i>P</i> ^a	ATM series (excluding MS in HBOC families) (N=31)	<i>P</i> ^a
Histological subtype							
Ductal carcinoma	434	30	Ref.	27	Ref.	25	Ref.
Lobular carcinoma	54	2	0.55	2	0.69	2	0.78
Others	28	3	0.43	2	0.78	3	0.25
Unknown	0	1	-	1	-	1	-
Architecture							
1-2	126	0	Ref.	7	Ref.	6	Ref.
3	390	30	0.97	21	0.86	20	0.96
Unknown	0	6	-	4	-	5	-
Mitosis							

0-1	221	15	Ref.	14	Ref.	12	Ref.
2	103	9	0.58	8	0.70	9	0.31
3	191	6	0.10	6	0.13	5	0.15
Unknown	1	6	-	4	-	5	-

Anisokaryosis

1	30	2	Ref.	2	Ref.	2	Ref.
2	225	7	0.25	7	0.22	5	0.12
3	261	21	0.96	19	0.89	19	0.9
Unknown	0	6	-	4	-	5	-

Histological grade

I	88	4	Ref.	3	Ref.	4	Ref.
II	191	16	0.34	15	0.24	13	0.58
III	236	15	0.64	13	0.55	14	0.75
Unknown	1	1	-	1	-	-	-

Tumour size (cm)

pT1 (<2)	329	23	Ref.	20	Ref.	19	Ref.
pT2 (2 - 5)	166	8	0.37	7	0.41	7	0.48
pT3 (>5)	14	2	0.37	2	0.30	2	0.26
pT4	7	0	-	0	-	0	-
Unknown	0	3	-	3	-	3	-

Pushing margins

Absent	461	19	Ref.	17	Ref.	15	Ref.
Present	49	2	0.98	2	0.93	2	0.78
Unknown	6	15	-	13	-	14	-

Emboles

Absent	315	13	Ref.	11	Ref.	13	Ref.
Present	199	14	0.08	14	0.04	11	0.32

Unknown	2	9	-	7	-	7	-
N stage							
pN0	286	18	Ref.	15	Ref.	15	Ref.
pN1	151	10	0.80	9	0.66	9	0.55
pN2	59	1	0.23	1	0.32	1	0.35
pN3	16	1	0.96	1	0.81	1	0.76
pNx	5	6	-	6	-	5	-
Estrogen receptor							
Positive	307	34	Ref.	31	Ref.	29	Ref.
Negative	209	1	0.003	1	0.004	1	0.005
Unknown	0	1	-	-	-	1	-
Progesterone receptor							
Positive	278	26	Ref.	23	Ref.	23	Ref.

Negative	236	8	0.03	8	0.07	6	0.02
Unknown	2	2	-	1	-	2	-
HER2							
Negative	433	25	Ref.	23	Ref.	21	Ref.
Positive	83	6	0.77	5	0.97	6	0.53
Unknown	0	5	-	4	-	4	-
Ki67							
<20%	173	11	Ref.	10	Ref.	8	Ref.
≥20%	336	18	0.55	17	0.59	17	0.97
Unknown	7	7	-	5	-	6	-
Molecular subtype							
TNBC	142	1	Ref.	1	Ref.	1	Ref.
HER2	66	0	N/A	0	N/A	0	N/A

Luminal A	180	10	0.06	9	0.08	7	0.13
Luminal B	111	13	0.009	12	0.010	12	0.010
Luminal B / HER2	17	4	0.005	4	0.005	4	0.005
Unknown	0	8	-	6	-	7	-

1 ^a P-value adjusted for gender and age at diagnosis

2 TNBC, triple-negative breast cancer

1 **Table 3.** Copy number losses recurrently observed in the 23 *ATM*-associated breast tumours.

2

Locus	Associated morphology	All ATM tumours (N=23)			ATM tumours with proven biallelic inactivation of ATM (N=16)			ATM tumours sequenced by WGS (N=4)		
		Cytogenetic band	Number of genes	Cancer genes*	Cytogenetic band	Number of genes	Cancer genes*	Cytogenetic band	Number of genes	Cancer genes*
6q	20% of ATM tumours							6q23.3-q27	190	<i>ARID1B, ECT2L, ESR1, EZR, FGFR10P, QKI, TNFAIP3</i>
8p	Breast cancer	8p21.3	6 ^d	-	8p23.3-p12	216	<i>WRN, NRG1</i>	8p21.3	6 ^d	-
11q	ATM tumours	11q22	2	<i>ATM, ZBTB16</i>	11q21-q25	276	<i>ATM, ZBTB16, DDX10, POU2AF1, SDHD, ZBTB16, PAFAH1B2,</i>	11q21-q24.2	236	<i>ATM, ZBTB16, DDX10, POU2AF1, SDHD, ZBTB16, PAFAH1B2,</i>

							<i>PDCK7, KMT2A,</i>			<i>PDCK7, KMT2A,</i>
							<i>MAML2, DDX6,</i>			<i>MAML2, DDX6,</i>
							<i>BCL9L, CBL,</i>			<i>BCL9L, CBL</i>
							<i>ARHGEF12, KCNJ5</i>			
13q	70% ATM tumours	13q14.11- q14.3	90	<i>LCP1, RB1</i>	13q13.3- q32.3	320	<i>LHFP, FOXO1,</i> <i>LCP1, RB1</i>	13q13.3- q32.3	320	<i>LHFP, FOXO1,</i> <i>LCP1, RB1</i>
16q	Luminal tumours	16q13-q24.3	364	<i>HERPUD1,</i> <i>CDH11, CFBF,</i> <i>CTCF, CDH1,</i> <i>ZFHX3, MAF,</i> <i>CBFA2T3,</i> <i>FANCA</i>	16q22.1	1 ^c	-	16q22.1	1 ^c	-
17p	Luminal tumours	17p13.3	46	<i>YWHAE</i>		-	-	17p13.3	46	<i>YWHAE</i>
17p	Luminal tumours	17p13.2-p12	166	<i>USP6,</i> <i>RABEP1,</i>		-	-			

								<i>TP53, GAS7,</i>
								<i>MAP2K4</i>
19p	40% of ATM tumours							19p13.3- p13.2 <i>FSTL3, GNA11,</i> <i>MAP2K2, MLLT1,</i> <i>SH3GL1, STK11,</i> <i>TCF3</i>
21p	70% of ATM tumours	21p11.2- p11.1	2 ^a	-	21p11.2- p11.1	2 ^a	-	
22q	Luminal B tumours, 70% ATM tumours	22q11.23	2 ^b		22q11.23	3 ^b		<i>GSTT1</i>
	Luminal B tumours, 70% ATM tumours	-	-	-	22q12.3	199		<i>APOBEC3B, MKL1,</i> <i>EP300</i>

1 * as reported in the COSMIC database.

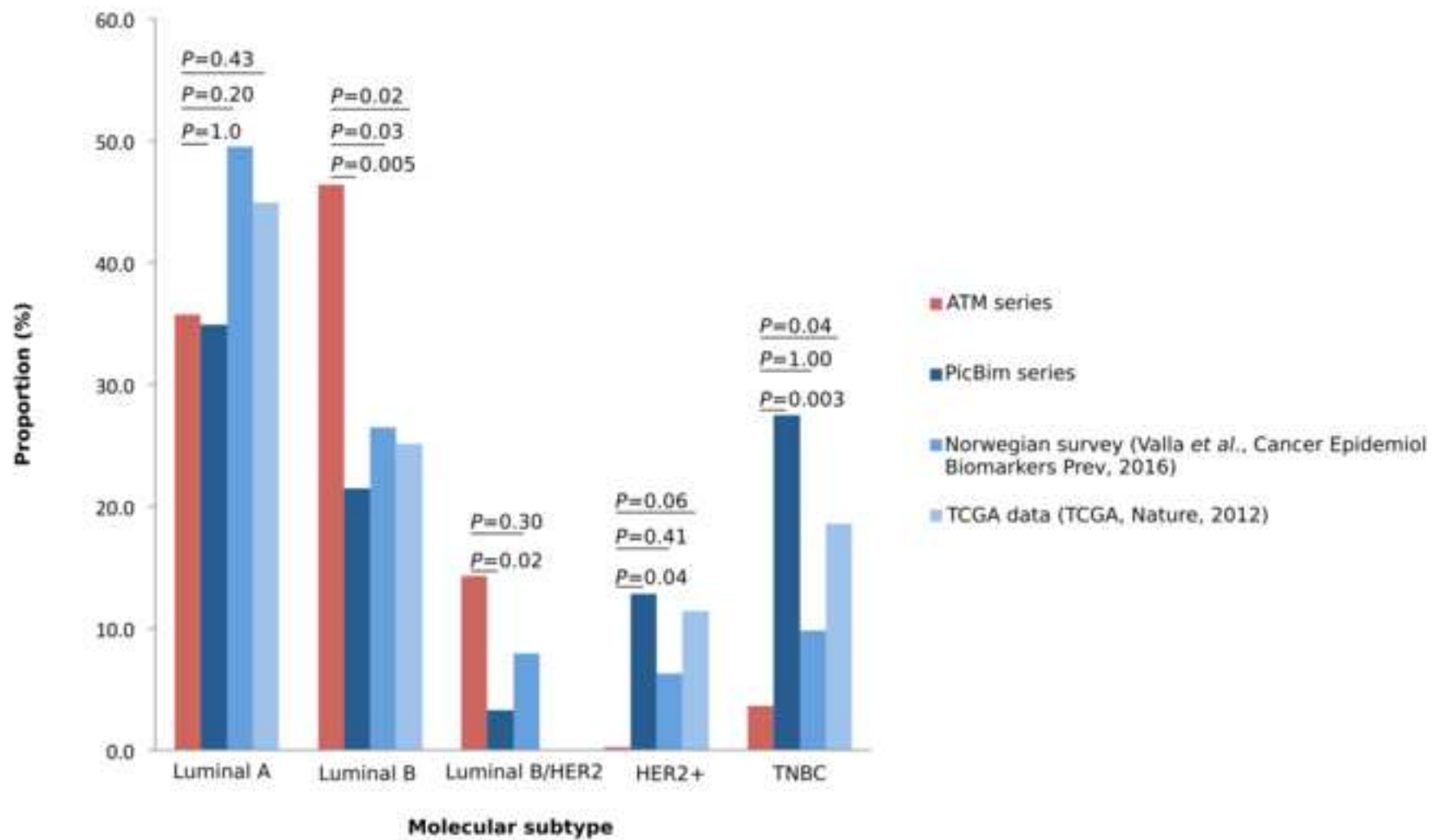
1 ^aThis region contains TEKT4P2, TPTE and 4 miRNAs (MIR3648-1, MIR3648-2, MIR3687-1, MIR3687-2).

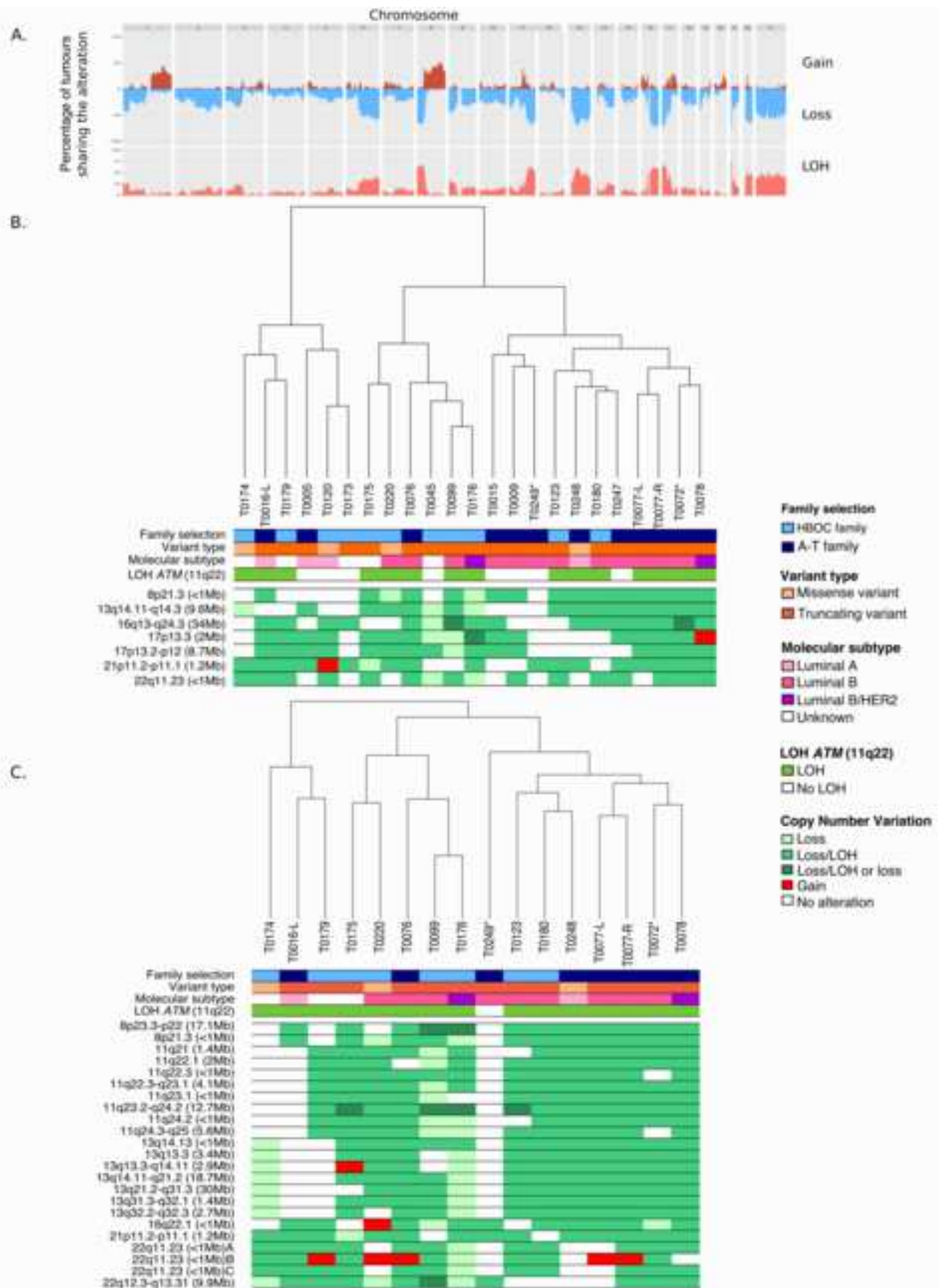
2 ^bThis region contains the pseudogenes *GSTTP1* and *GSTTP2*.

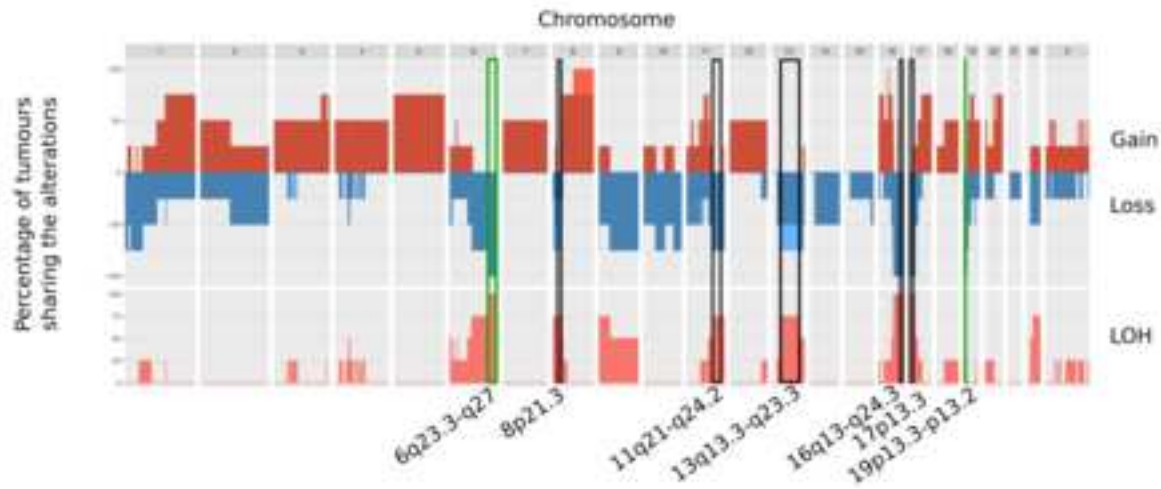
3 ^cThis region contains WWP2 and the miRNA MIR140.

4 ^dThis region contains *PPP3CC*, *SORBS3*, *PDLIM2*, *BIN3*, *BIN3-IT1*, *EGR3*.

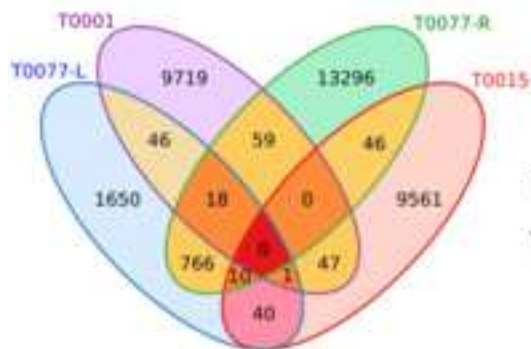
5



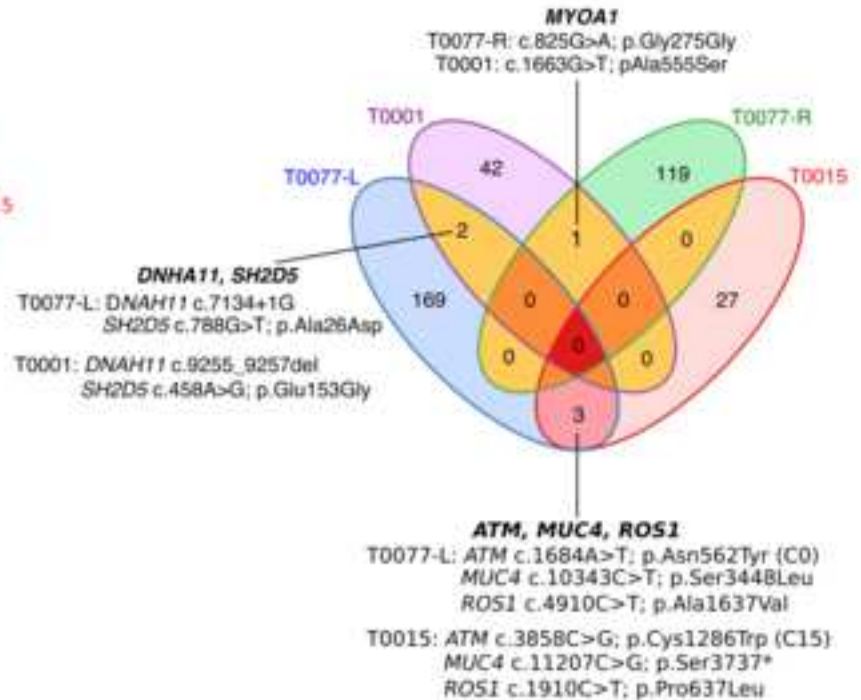




B.



C.



5. DISCUSSION

L'un des défis de cette étude aura été la collecte d'un nombre suffisant de tumeurs. La législation autorise les laboratoires d'anatomopathologie à détruire le matériel biologique conservé depuis plus de 10 ans, aussi nous n'avons pu exploiter/analyser le matériel tumoral d'une partie importante des porteuses *ATM* identifiées dans les études CoF-AT et GENESIS. De plus, les laboratoires français ayant pendant des années utilisé le liquide de Bouin comme fixateur, les analyses moléculaires sur ces tumeurs ont d'office été exclues à cause des dégradations de l'ADN engendrées par cet agent chimique.

L'ensemble des résultats obtenus dans cette partie de la thèse montrent que les tumeurs du sein *ATM* ne sont pas triple-négatives et ne présentent pas la signature HRD, suggérant que des mécanismes de tumorigenèse différents sont impliqués dans développement des tumeurs *ATM* et *BRCA1*. En revanche, le sous-type moléculaire seul ne permet pas de distinguer les tumeurs du sein *ATM* de celles développées par les porteurs d'une mutation dans un autre gène de prédisposition au cancer, comme les tumeurs *BRCA2* ou *CHEK2*, elles aussi majoritairement lumineales.

Néanmoins, des caractéristiques spécifiques aux tumeurs du sein *ATM* semblent émerger de ces travaux. En effet, les tumeurs du sein *ATM* pourraient se distinguer des tumeurs des cas sporadiques par la présence d'une tétraploïdie, une perte d'hétérozygotie au locus 11q22-23 contenant *ATM* et à une perte du nombre de copies au niveau des régions 13q, 21p et 22q. La perte de l'allèle normal au locus *ATM* observée dans 70% des tumeurs de patients HetAT analysées confirme le rôle d'*ATM* dans le développement du cancer du sein et pourrait expliquer la tétraploïdie dans ces tumeurs. En effet, l'absence de protéines *ATM* fonctionnelles empêcherait les points de contrôle du cycle cellulaire de se faire normalement entraînant des fusions de cellules ou des endo-réplifications qui conduiraient à la formation de cellules tétraploïdes. D'autre part, les régions identifiées en perte ou en LOH devront être validées par l'analyse de l'expression des gènes localisés dans ces régions.

L'une des limites de notre étude est le peu de matériel tumoral congelé disponible pour l'analyse du génome tumoral (WGS pour « *whole-genome sequencing* »). Dû au petit nombre de tumeurs analysées, peu de résultats sont ressortis de cette analyse. L'analyse CNV des quatre génomes réplique les résultats obtenus à partir de l'analyse Oncoscan et ne met pas en évidence de nouvelles régions. Par ailleurs, l'analyse des SNVs sur l'ensemble du génome montre que les tumeurs partagent peu de variants. Ce résultat se reflète également au niveau des parties codantes du génome, puisque seuls six gènes ont été identifiés comme altérés dans au moins deux tumeurs (*MUC4*, *MYO1A*, *ATM*, *SH2D5*, *DNHA11*). Seuls les gènes *MUC4* et *DNHA11* présentent un variant délétère, potentiellement driver dans la tumorigenèse. On notera qu'une diminution de l'expression du gène *MUC4* est associée à la progression tumorale et à une augmentation des cellules immunitaires TCD8 et des

cellules NK (Natural Killer) dans l'environnement tumoral. Ce dernier point pourrait expliquer la présence d'un infiltrat inflammatoire modéré dans les tumeurs T0077-L et T0015 présentant une altération de ce gène.

Pour aller plus loin, en collaboration avec Elodie Girard, ingénieure bioinformaticienne travaillant sur la plateforme de bioinformatique de l'U900, nous avons recherché si l'une des signatures mutationnelles mise en évidence par Nik-Zainal et *coll.* était identifiable dans ces quatre tumeurs³⁴ (Tableau 17 et Figure 28). La définition de ces signatures ne nous a pas permis d'apporter une information supplémentaire claire concernant la caractérisation des tumeurs *ATM*. En effet, les signatures mutationnelles obtenues à partir des quatre tumeurs sont principalement associées à l'âge et au tabac, ce qui concorde avec l'âge au diagnostic et au passé de fumeuses pour les patientes T0077 et T0001 (Tableau 17). On notera néanmoins l'absence d'association avec la signature mutationnelle associée à une déficience de la réparation de l'ADN par recombinaison homologue, ce qui est en adéquation avec les résultats d'OncoScan ne montrant pas de profil LST dans les tumeurs *ATM*.

Tableau 17 : Signatures mutationnelles identifiées parmi les 4 génomes à partir du catalogue de signature publié par MR. Stratton.

Signature	Description	T0077-L	T0077-R	T0015	T0001
Signature 1A	Déamination 5-méthyl cytosine, âge	0.78	0.47	0.59	0.47
Signature 4	Tabac	0	0.39	0.26	0.1
Signature 15	Réparation des bases non-appariées déficiente	0.15	0	0	0
Signature 3	Réparation par recombinaison homologue déficiente	0.07	0	0	0.1
Signature 1B	Déamination 5-méthyl cytosine, âge	0	0	0	0
Signature 12	Inconnue	0	0	0	0.07
Signature 9	Hypermutation du gène de l'immunoglobine	0	0	0	0.07

Les chiffres rapportés pour chaque patient pour les différentes signatures correspondent à la proportion de mutations dans le génome associées à chaque signature. Par exemple, 78% des mutations de la tumeur T0077-L sont associées à la signature 1A liée à l'âge.

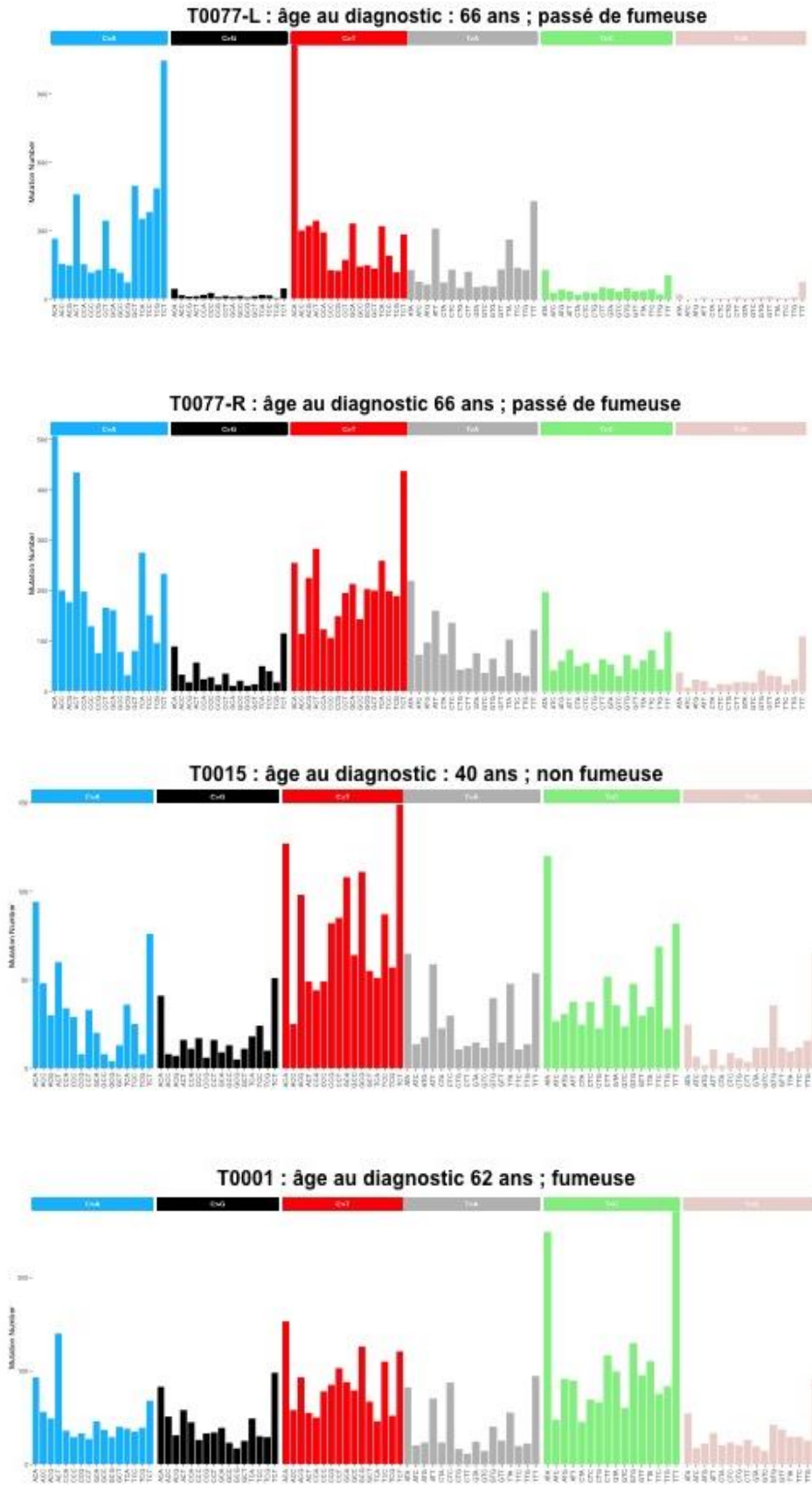


Figure 28 : Profils mutationnels des 4 génomes à partir du catalogue de MR. Stratton.

La majorité des tumeurs ATM étudiées étant conservées en FFPE, le WGS à partir de ce type de

matériel n'est pas réalisable. Afin de compléter la caractérisation génomique des tumeurs *ATM* nous avons initié en collaboration avec la plateforme ICGEx de l'Institut Curie (Institut Curie Génomique d'Excellence) un projet pilote pour évaluer la faisabilité de séquencer l'exome (WES pour « *whole-exome sequencing* ») de tumeurs conservées en FFPE.

Le projet pilote a été réalisé sur quatre paires d'ADN normal/ADN tumoral à partir de deux protocoles de préparation des banques : les protocoles MedExome (Roche) et SureSelect (Agilent), couvrant respectivement 46Mb et 54Mb du génome. Afin de déterminer lequel de ces protocoles produit les données les plus fiables, nous avons utilisé comme contrôles qualité les données de séquençage du génome pour les patients T0077-R et T0015, et les données d'OncoScan pour les patients T0009 et T0076 (Tableau 18).

Tableau 18 : Matériel disponible et analyses réalisées sur les paires d'ADN normal/ADN tumoral pour 4 échantillons.

Sample ID	WGS	OncoScan	SureSelect (Agilent)	MedExome (Roche)
Patient T0077-sang	✓			
Tumeur T0077R-congelé	✓		✓	✓
Tumeur T0077-FFPE		✓	✓	✓
Patient T0015-sang	✓			
Tumeur T0015-congelé	✓		✓	
Tumeur T0015-FFPE		✓	✓	✓
Patient T0009-sang			✓	✓
Tumeur T0009-FFPE		✓	✓	✓
Patient T0076-sang			✓	✓
Tumeur T0076-FFPE		✓	✓	✓

L'ensemble des données obtenues à partir du séquençage du génome pour les patients T0077 et T0015 et l'analyse OncoScan pour les patients T0009 et T0076 ont permis d'établir des contrôles qualités. Les SNPs ayant une MAF>1% identifiés par WGS ont été comparés aux SNPs identifiés dans les 2 kits sur l'ADN congelé ou l'ADN FFPE. Pour les patients T0009 et T0076 les SNPs obtenus à partir des données d'OncoScan ont servi de contrôles.

La comparaison des données obtenues à partir des deux protocoles a été réalisée sur les régions communes séquencées, c'est à dire sur 39Mb. Globalement, les données de WES obtenues à partir des deux protocoles montrent une couverture similaire pour les différents types de matériel (Tableau 19). Les contrôles qualité montrent que pour les deux protocoles, plus de 95% des SNPs exoniques identifiés par WGS ou génotypés avec la OncoScan sont retrouvés dans les données SureSelect et MedExome. Environ 50% des SNPs manquant dans les deux protocoles sont liés à une trop faible couverture des régions concernées. Les variants somatiques de type SNV et indels identifiés par WES ont été comparés à ceux identifiés par WGS pour les patients T0077 et T0015, et les résultats montrent un nombre plus élevés de variants identifiés uniquement avec le protocole MedExome, notamment à partir de matériel FFPE (Figure 29). Pour éviter un taux trop important de faux-positifs nous avons donc choisi de séquencer l'ensemble du matériel disponible avec le protocole SureSelect.

Le séquençage d'une deuxième série de huit tumeurs est actuellement en cours.

Tableau 19 : Profondeur moyenne de séquençage selon le type de matériel et le protocole de préparation des banques utilisées.

Matériel	Protocole	Profondeur moyenne (X)
Blood	MedExome	279
Blood	SureSelect	148
Frozen	MedExome	281
Frozen	SureSelect	241
FFPE	MedExome	207
FFPE	SureSelect	160

Les échantillons séquencés avec les protocoles SureSelect ont une plus faible couverture moyenne que les échantillons séquencés avec MedExome. Ces résultats sont probablement dû à un biais de protocole car les « runs » de séquençage ont été réalisés à partir de 8 banques SureSelect versus 6 banques MedExome, ce qui a favorisé l'amplification des banques MedExome.

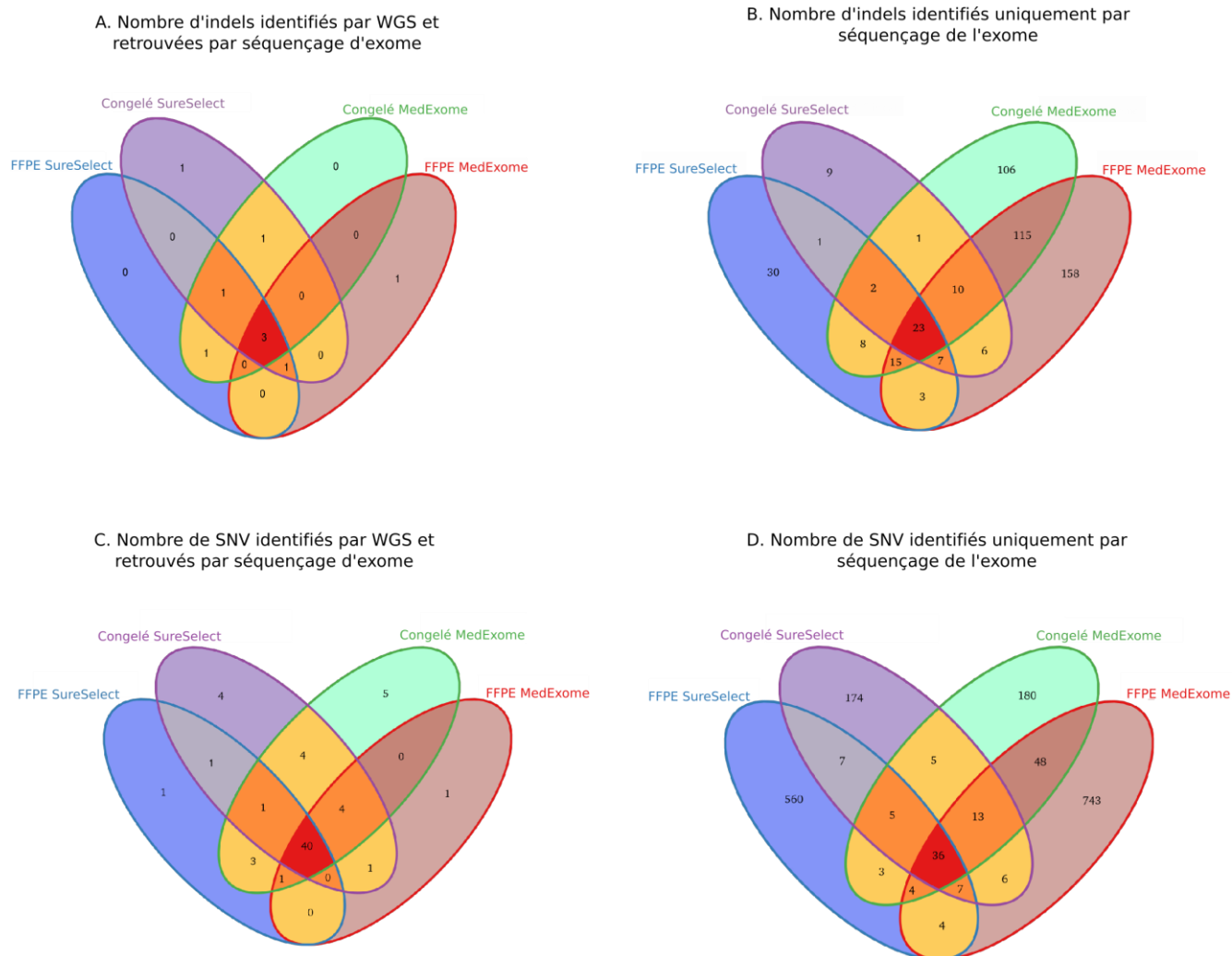


Figure 29 : Diagramme de Venn décrivant le nombre de variants exoniques somatiques communs identifiés par WGS et par WES, et le nombre de variants identifiés uniquement par WES en fonction du matériel chez les patients T0077 et T0015.

A. Parmi les 13 InDels identifiées par WGS, 9 sont retrouvées dans au moins une tumeur séquencée par WES **B.** 494 InDels ont été trouvées par WGS, 76% d'entre elles sont trouvées avec le protocole MedExome. **C.** Parmi les 121 SNV identifiées par WGS, 66 sont retrouvées dans au moins une tumeur séquencée par WES. **D.** 1795 SNV ont été mis en évidence par WES, 54% d'entre elles sont uniquement trouvées avec le protocole MedExome.

Partie IV : Discussion & perspectives

1. Nouvelles découvertes associées à ce projet de thèse

L'objectif de mes travaux était d'identifier des facteurs génétiques constitutionnels pouvant moduler le risque de cancer du sein chez les porteurs d'une mutation constitutionnelle d'*ATM* et de décrire la morphologie et les caractéristiques moléculaires des tumeurs mammaires développées par les porteurs d'une ou deux copies altérées du gène.

Ce projet a été réalisé à partir de la ressource biologique collectée dans le cadre des études rétrospectives et prospectives sur les familles A-T, CoF-AT et Retro-AT. Lors de l'initiation du projet de thèse, seules les mutations familiales d'*ATM* avaient été recherchées chez les participantes mais aucune autres variations génétiques n'avait encore été recherchée dans ces familles. Le séquençage d'un panel de 55 gènes chez les femmes et patients A-T atteints d'un cancer a permis de mieux caractériser les mutations ségrégant dans ces familles.

Dans la première partie de la thèse, je me suis concentrée sur l'analyse de la longueur des télomères et le génotypage de SNPs associés à la longueur des télomères et/ou au cancer du sein. Les résultats montrent de façon surprenante que les individus HetAT ont des télomères plus longs que leurs apparentés non-HetAT. Les connaissances sur les mécanismes de régulation de la longueur des télomères chez les individus HetAT sont à élargir pour comprendre cette observation et identifier des biomarqueurs potentiels en lien avec les fonctions d'*ATM*. Bien que nos résultats ne montrent pas d'association entre la longueur des télomères et le risque de cancer dans les familles A-T, les porteurs HetAT ayant des télomères courts semblent néanmoins plus à risque de développer un cancer que les non porteurs. Si ces résultats étaient confirmés dans de plus larges études, ce facteur génétique constitutionnel pourrait être pris en compte dans un model d'estimation des risques pour les individus HetAT.

D'autre part, sur les 21 SNPs associés à la longueur des télomères et/ou au cancer du sein et génotypés dans la population d'étude, seul le SNP rs9257445 dans le gène *ZNF311* a été associé à la longueur des télomères chez les individus HetAT. Ce SNP pourrait donc servir de proxy pour déterminer la longueur des télomères chez les individus HetAT, le génotypage par TaqMan étant une méthode plus simple et moins couteuse que la méthode q-PCR permettant d'estimer la longueur relative des télomères.

Par ailleurs, le SNP rs6060627 dans le gène *BCL2L1* est associé à une diminution du risque de cancer chez les femmes HetAT et non-HetAT de l'étude. Bien que l'ensemble des SNPs testés dans le cadre de cette thèse ne correspondent pas à une liste exhaustive de facteurs génétiques pouvant moduler le risque de cancer, nos travaux soulignent que, comme déjà mis en évidence chez les porteurs *BRCA1* et *BRCA2*^{126-128, 287, 295}, des facteurs génétiques peuvent également moduler le risque de cancer

du sein chez les porteurs de mutation dans un gène de prédisposition conférant un risque de cancer plus modéré. Les SNPs aujourd'hui considérés dans les modèles de prédiction de risque (PRS) en population générale (études du consortium BCAC) ou chez les porteurs d'une mutation de *BRCA1* ou *BRCA2* (études du consortium CIMBA) pourraient donc moduler de façon significative le risque de cancer des sujets apparentés à un enfant atteint d'A-T et seraient à tester dans les familles *ATM*.

Au niveau somatique, l'analyse des caractéristiques morphologiques et moléculaires des tumeurs du sein développées par les porteurs d'une ou deux copies mutées d'*ATM* a permis de mettre en évidence une signature *ATM* associée au sous-type moléculaire luminal B, à la tétraploïdie des cellules tumorales, à l'inactivation d'*ATM* dans les tumeurs due à la perte d'hétérozygotie de l'allèle normal ainsi que des pertes du nombre de copies aux loci 13q, 21p et 22q. Ce travail a permis d'enrichir les connaissances sur les tumeurs du sein *ATM* et également de montrer que ces tumeurs ne ressemblent pas aux tumeurs *BRCA1* majoritairement triple-négatives et caractérisées par la signature HRD au niveau moléculaire. Bien que la signature *ATM* mise en évidence dans ces travaux ne permette pas de prédire la sensibilité aux traitement ciblant les déficiences de la réparation, l'analyse des CNV sur l'ensemble du génome a permis de mettre en évidence des pertes au niveau de quatre gènes d'intérêt : *RB1* impliqué dans le rétinoblastome et associé à une faible réponse à l'hormonothérapie, *LCP1* considéré comme un biomarqueur de la sévérité de la tumeur, *TPTE* appartenant à la famille du gène *PTEN* et enfin les délétions de *GSTT1*. De telles délétions ont été associées à une augmentation du risque de rechute de cancer du sein (OR=4,25 ; p=0,04)³⁵⁹. Afin de déterminer si l'expression de ces gènes est diminuée à cause des pertes chromosomiques observées nous avons initié au laboratoire une étude pilote sur l'analyse du transcriptome des tumeurs du sein décrites dans l'Article 2.

2. Forces et limites

L'une des limites de cette étude est le petit nombre d'échantillons disponibles dû à la rareté du syndrome de l'A-T et donc à la difficulté d'identifier des individus HetAT. Néanmoins, la ressource de CoF-AT est une ressource unique au monde, bien caractérisée, sans laquelle ce projet n'aurait pas pu être mené. En raison du petit nombre de participant, certaines analyses n'ont pas pu être réalisées, comme par exemple l'association entre les SNPs associés au risque de cancer du sein en fonction du sous-type moléculaire mis en évidence lors de la relecture anatomopathologique. De même, ce critère de sélection des tumeurs n'a pu être pris en compte dans l'analyse de la longueur des télomères chez les sujets HetAT. Nous n'avons pas non plus évalué la contribution des SNPs et de la longueur des télomères dans le développement des cancers à une autre localisation que le sein.

En revanche, l'une des forces de notre étude sur les télomères réside dans le fait que les femmes participant à CoF-AT étaient indemnes de tous cancers à l'inclusion dans l'étude ; l'analyse de la longueur des télomères réalisée sur l'ADN des lymphocytes a donc pu être réalisée sans introduire de biais liés aux traitements anti-tumoraux.

Le recueil de données épidémiologiques, cliniques et biologiques des porteuses d'une mutation d'*ATM* est toujours en cours. Tout au long de ces quatre années de thèse, j'ai mené mes travaux en étroite collaboration avec les attachées de recherche en épidémiologie de la PIGE au sein de l'équipe de Nadine Andrieu et ai pu bénéficier de leur expérience pour recueillir des données de haute qualité. Afin d'éviter les biais de déclarations par exemple, les questionnaires épidémiologiques sont revus systématiquement, et complétées si besoin par Eve Cavaciuti et Dorothée Le Gal lors d'un rendez-vous téléphonique avec la participante. J'ai également pu exploiter les données médicales des participantes pour lesquelles le diagnostic avait été validé auprès du médecin traitant.

Au début du projet sur la caractérisation tumorale, toutes les tumeurs recensées n'avaient pas encore été centralisées dans notre laboratoire. Grâce aux interactions avec l'équipe d'investigation de la PIGE nous avons pu mettre en place un système de collecte des tumeurs automatisées entre le centre coordonnateur (U900), le service d'anatomopathologie de l'Institut Curie et le centre de ressource biologique. De plus, ces interactions ont abouti à la mise en place d'une base de gestion centralisant les données de la relecture histologiques des tumeurs du sein. Par ailleurs, les traitements néo-adjuvants pouvant avoir un impact sur la morphologie des tumeurs, nous sommes assurés qu'aucun des patients n'avaient reçu ce type de traitement.

L'Institut Curie est un hôpital possédant une grande expertise dans le suivi et le traitement des patients atteints de cancer du sein. Pour ce projet de recherche nous avons pu bénéficier de l'expertise des anatomopathologistes Anne Vincent-Salomon et Guillaume Bataillon qui ont réalisé la relecture histologique à l'aveugle de l'ensemble des tumeurs analysées dans l'article rapportant les résultats sur la caractérisation histologique et moléculaire des tumeurs du sein *ATM*. Cette relecture systématique par deux anatomopathologistes de référence avait pour but d'empêcher les biais dans la façon d'attribuer les scores aux différentes caractéristiques tumorales.

L'une des limites du projet sur la caractérisation des tumeurs aura été le faible nombre de tumeurs congelées disponibles, ce qui n'aura pas permis de mettre en évidence des mutations drivers qui auraient pu constituer de nouvelles cibles thérapeutiques ; cependant grâce à l'implication de l'équipe dans des consortia internationaux tels que ENIGMA (*Evidence-based Network for the Interpretation of Germline Mutant Alleles*; <https://enigmaconsortium.org/>) nous avons entrepris d'étendre l'analyse des tumeurs *ATM* à d'autres collections et avons notamment travaillé avec l'équipe de Georgia Chenevix-Trench du centre anti-cancer Peter MacCallum à Melbourne en Australie (Article 2).

3. Impact et futur

Les enjeux de la recherche sur le cancer sont multiples. En effet, une partie du risque génétique lié au cancer du sein reste actuellement inexpliquée. Les études pangénomiques comme l'OncoArray¹³¹, ou les projet séquençage à haut débit (NGS) dans de larges études ont pour but d'identifier de nouveaux loci ou gènes de prédisposition au cancer du sein. Par ailleurs, l'introduction des tests multi-gènes pour le diagnostic moléculaire des familles à haut risque de cancer du sein pose la question de l'utilité clinique des gènes de prédisposition de risque intermédiaire tels que *ATM*, *CHEK2*, *BRIP1* ou des « nouveaux » gènes comme *XRCC2* ou *RINT1*. En effet, de plus en plus de porteurs de variants dans ces gènes sont identifiés par les laboratoires de diagnostic ou dans des projets de recherche d'épidémiologie moléculaire dans lesquelles de larges séries cas-témoins sont analysées. Cependant les données familiales sont encore trop peu nombreuses pour pouvoir estimer précisément le risque absolu de cancer du sein pour les porteurs d'un variant prédit comme délétère dans ces gènes, ce qui empêche d'élaborer des recommandations pour le dépistage et le suivi des familles.

Depuis 2006, le rôle d'*ATM* dans la prédisposition génétique au cancer du sein n'est plus à démontrer et les mutations de ce gène sont associées à un risque intermédiaire¹³⁵. De plus en plus de données de séquençage d'exomes dans des familles BRCAx sont publiées et estiment une fréquence de mutation d'*ATM* d'au moins 2% dans ces familles à haut-risque, suggérant une prévalence des mutations d'*ATM* plus importante qu'initialement estimée. D'autre part, on estime que 0,5% à 1% de la population générale serait porteuse d'une mutation délétère d'*ATM*. Afin de mieux estimer les risques absolus de cancer pour les porteurs de mutation dans *ATM*, ainsi que pour les porteurs de variants délétères dans des gènes de prédisposition conférant un risque intermédiaire ou dans des « nouveaux » gènes, l'étude TUMOSPEC a été initiée par le Dr Olivier Caron et notre équipe en collaboration avec l'ensemble des partenaires du Groupe Génétique et Cancer d'UNICANCER (consultations d'oncogénétiques et laboratoires de diagnostic moléculaire). Les patients porteurs d'une mutation d'*ATM*, ou d'un variant faux-sens prédit comme délétère identifiés via TUMOSPEC seront invités en consultation d'oncogénétique à participer à l'étude de suivi CoF-AT2. Dans cette extension de l'étude CoF-AT, tous les membres de la famille seront invités à participer, hommes et femmes, porteurs et non porteurs. L'étude CoF-AT2 permettra d'étudier l'étiologie du cancer du sein chez plus de femmes HetAT mais également d'étudier d'autres types de cancer. En effet, d'après les données familiales de CoF-AT les porteurs HetAT, hommes et femmes, présentent un risque accru du cancer du pancréas (SIR=4,95, 95% IC :3,04 - 7,61), et de lymphome ou de leucémie chez les apparentés ayant une probabilité d'être hétérozygote pour *ATM* (SIR=5,33, 95% IC : 1,43 - 13,7) (Lecarpentier J, Andrieu N, communication personnelle).

La pénétrance incomplète des mutations d'*ATM* soulève la question de la présence d'autres facteurs, génétiques ou environnementaux qui pourraient modifier le risque de cancer du sein des individus HetAT. Les résultats mis en évidence par le *Consortium of Investigators of Modifiers of BRCA1/2* (CIMBA) montrent que le risque de cancer du sein et/ou de l'ovaire chez les porteurs d'une mutation de *BRCA1* ou *BRCA2* est modulé par la présence de SNPs associés au cancer du sein dans la population générale^{126, 128, 287, 295}. Dans notre étude sur les sujets HetAT, la mise en évidence d'un SNP associé à la longueur des télomères et d'un deuxième SNP associé au cancer du sein montre que l'analyse de ces variants fréquents n'est pas uniquement à considérer chez les porteurs de mutations à forte pénétrance mais également chez les porteurs de mutation à pénétrance intermédiaire.

D'autre part, des études internationales s'intéressent également au rôle des expositions environnementales sur le risque de cancer. Le consortium *International BRCA1/2 Carrier Cohort Study* (IBCCS) a montré chez les porteurs d'une mutation *BRCA1* et *BRCA2* que l'exposition aux radiations ou le nombre de grossesses menées à terme étaient des facteurs modificateurs du risque de cancer du sein chez ces porteurs^{360, 361}. Grâce aux données de suivi de CoF-AT et CoF-AT2, de telles analyses seront possibles pour les porteurs de variants d'*ATM*.

L'un des autres enjeux actuels de la recherche est la classification des variants. En effet, de nombreux variants localisés dans des gènes de prédisposition au cancer du sein autre que *BRCA1* et *BRCA2* n'ont pas d'impact clinique démontré et sont donc bien souvent classés comme variant de signification clinique inconnue (ou « VSI »). Une étude sur 660 femmes non porteuses de mutations *BRCA1* ou *BRCA2* a révélé que pour seulement 2% d'entre elles le séquençage d'un panel de gènes aura eu une utilité clinique³⁶². Néanmoins, pour aider à la classification des VSI plusieurs outils sont mis en place pour faciliter l'interprétation de ces variants par les généticiens et les chercheurs. Pour les gènes *BRCA1* et *BRCA2*, ces outils consistent en la base française UMD-*BRCA1/BRCA2* (<http://www.umd.be/>), la classification de l'IARC reposant sur la co-ségrégation du variant, l'histoire personnelle et familiale, la pathologie de la tumeur et la co-occurrence du variant en trans et la base ClinVar pour laquelle les variants de *BRCA1* et *BRCA2* notamment sont classés par un panel d'expert. Pour le gène *ATM*, la classification des variants identifiés dans le contexte d'une prédisposition familiale au cancer du sein ou de l'ovaire, ou d'un autre cancer mais hors contexte A-T reste problématique. L'une des stratégies de classifications des variants faux-sens ou des variants classés VSI serait l'utilisation de la signature tumorale *ATM*. Les résultats rapportés dans l'Article 2 seront confirmés dans une deuxième série de tumeurs collectées à partir des collaborations mises en place avec des partenaires impliqués dans le consortium ENIGMA lors de cette thèse, notamment avec les laboratoires de diagnostic moléculaire de Chicago (Université de Chicago, F. Olopade, S. Nielsen),

d'Athènes (National Centre for Scientific Research Demokritos, F. Fostira), l'hôpital Sírio-Libanês (São Paulo, Brazil, E. Santana dos Santos) et avec l'étude Malaysian breast cancer genetics (S.Teo). Si la signature moléculaire est validée dans cette seconde série de tumeurs *ATM* provenant de patients vus en consultation de génétique, son utilisation comme outil de classification sera alors envisageable pour déterminer si les variants de types faux-sens sont les variants causaux de la maladie, selon la présence ou non de la signature *ATM*.

D'autre part, cette signature permettrait également d'identifier des femmes porteuses d'une mutation d'*ATM* hors contexte A-T et également hors contexte d'une prédisposition familiale.

La caractérisation des variants ainsi qu'une meilleure estimation des risques permettront *in fine* de stratifier les femmes à risque pour mieux adapter leur dépistage et leur suivi. Les femmes à risque élevé seraient suivies plus précocement et de façon plus fréquente ce qui à long terme réduirait les effets secondaires des traitements et augmenterait la qualité de vie des patientes. Inversement, une stratification des femmes à faible risque diminuerait les coûts de santé publique en réduisant le nombre de mammographies.

REFERENCES BIBLIOGRAPHIQUES

1. Sinn HP, Kreipe H. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel)* 2013;**8**: 149-54.
2. Malhotra GK, Zhao X, Band H, Band V. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther* 2010;**10**: 955-60.
3. Russnes HG, Lingjaerde OC, Borresen-Dale AL, Caldas C. Breast cancer molecular stratification- from intrinsic subtypes to integrative clusters. *The American journal of pathology* 2017.
4. Yang LY, Yang LP, Zhu B. Clinicopathological characteristics and survival outcomes of invasive lobular carcinoma in different races. *Oncotarget* 2017.
5. Hall FM. Papillary lesions of the breast. *Radiology* 2007;**243**: 300-1; author reply 1.
6. Rakha EA, Ahmed MA, Ellis IO. Papillary carcinoma of the breast: diagnostic agreement and management implications. *Histopathology* 2016;**69**: 862-70.
7. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clin Med Insights Pathol* 2015;**8**: 23-31.
8. Biswas T, Efirid JT, Prasad S, James SE, Walker PR, Zagar TM. Inflammatory TNBC Breast Cancer: Demography and Clinical Outcome in a Large Cohort of Patients With TNBC. *Clin Breast Cancer* 2016;**16**: 212-6.
9. Yamauchi H, Woodward WA, Valero V, Alvarez RH, Lucci A, Buchholz TA, Iwamoto T, Krishnamurthy S, Yang W, Reuben JM, Hortobagyi GN, Ueno NT. Inflammatory breast cancer: what we know and what we need to learn. *Oncologist* 2012;**17**: 891-9.
10. Vrieling C, van Werkhoven E, Maingon P, Poortmans P, Weltens C, Fourquet A, Schinagl D, Oei B, Rodenhuis CC, Horiot JC, Struikmans H, Van Limbergen E, et al. Prognostic Factors for Local Control in Breast Cancer After Long-term Follow-up in the EORTC Boost vs No Boost Trial: A Randomized Clinical Trial. *JAMA Oncol* 2017;**3**: 42-8.
11. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer* 1957;**11**: 359-77.
12. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991;**19**: 403-10.
13. Parise CA, Bauer KR, Brown MM, Caggiano V. Breast cancer subtypes as defined by the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) among women with invasive breast cancer in California, 1999-2004. *The breast journal* 2009;**15**: 593-602.
14. Ward HW. Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *Br Med J* 1973;**1**: 13-4.
15. Banerjee M, George J, Song EY, Roy A, Hryniuk W. Tree-based model for breast cancer prognostication. *J Clin Oncol* 2004;**22**: 2567-75.
16. MacGrogan G, de Mascarel I, Sierankowski G, Mauriac L, Debled M, Durand M, De Lara CT, Avril A, Picot V, Mathoulin-Pelissier S. Time for reappraisal of progesterone-receptor testing in breast cancer management. *J Clin Oncol* 2005;**23**: 2870-1; author reply 1.
17. Rakha EA, El-Sayed ME, Green AR, Paish EC, Powe DG, Gee J, Nicholson RI, Lee AH, Robertson JF, Ellis IO. Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. *J Clin Oncol* 2007;**25**: 4772-8.
18. Tal M, Wetzler M, Josefberg Z, Deutch A, Gutman M, Assaf D, Kris R, Shiloh Y, Givol D, Schlessinger J. Sporadic amplification of the HER2/neu protooncogene in adenocarcinomas of various tissues. *Cancer Res* 1988;**48**: 1517-20.
19. Baselga J, Norton L, Albanell J, Kim YM, Mendelsohn J. Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts. *Cancer Res* 1998;**58**: 2825-31.

20. Kast K, Schoffer O, Link T, Forberger A, Petzold A, Niedostatek A, Werner C, Klug SJ, Werner A, Gatzweiler A, Richter B, Baretton G, et al. Trastuzumab and survival of patients with metastatic breast cancer. *Arch Gynecol Obstet* 2017;**296**: 303-12.
21. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;**344**: 783-92.
22. von Minckwitz G, Rezai M, Fasching PA, Huober J, Tesch H, Bauerfeind I, Hilfrich J, Eidtmann H, Gerber B, Hanusch C, Blohmer JU, Costa SD, et al. Survival after adding capecitabine and trastuzumab to neoadjuvant anthracycline-taxane-based chemotherapy for primary breast cancer (GBG 40--GeparQuattro). *Ann Oncol* 2014;**25**: 81-9.
23. Chang J, Ormerod M, Powles TJ, Allred DC, Ashley SE, Dowsett M. Apoptosis and proliferation as predictors of chemotherapy response in patients with breast carcinoma. *Cancer* 2000;**89**: 2145-52.
24. Faneyte IF, Schrama JG, Peterse JL, Remijnse PL, Rodenhuis S, van de Vijver MJ. Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. *British journal of cancer* 2003;**88**: 406-12.
25. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol* 2005;**23**: 7212-20.
26. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, et al. Molecular portraits of human breast tumours. *Nature* 2000;**406**: 747-52.
27. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 2001;**98**: 10869-74.
28. Weigelt B, Horlings HM, Kreike B, Hayes MM, Hauptmann M, Wessels LF, de Jong D, Van de Vijver MJ, Van't Veer LJ, Peterse JL. Refinement of breast cancer classification by molecular characterization of histological special types. *J Pathol* 2008;**216**: 141-50.
29. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992;**258**: 818-21.
30. Hicks J, Muthuswamy L, Krasnitz A, Navin N, Riggs M, Grubor V, Esposito D, Alexander J, Troge J, Wigler M, Maner S, Lundin P, et al. High-resolution ROMA CGH and FISH analysis of aneuploid and diploid breast tumors. *Cold Spring Harb Symp Quant Biol* 2005;**70**: 51-63.
31. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**: 61-70.
32. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006;**10**: 529-41.
33. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;**486**: 346-52.
34. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016;**534**: 47-54.
35. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;**486**: 400-4.
36. Santarpia L, Bottai G, Kelly CM, Gyorffy B, Szekely B, Pusztai L. Deciphering and Targeting Oncogenic Mutations and Pathways in Breast Cancer. *Oncologist* 2016;**21**: 1063-78.
37. Kittaneh M, Montero AJ, Gluck S. Molecular profiling for breast cancer: a comprehensive review. *Biomark Cancer* 2013;**5**: 61-70.

38. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America* 2003;**100**: 10393-8.
39. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America* 2003;**100**: 8418-23.
40. Habashy HO, Powe DG, Abdel-Fatah TM, Gee JM, Nicholson RI, Green AR, Rakha EA, Ellis IO. A review of the biological and clinical characteristics of luminal-like oestrogen receptor-positive breast cancer. *Histopathology* 2012;**60**: 854-63.
41. Han W, Han MR, Kang JJ, Bae JY, Lee JH, Bae YJ, Lee JE, Shin HJ, Hwang KT, Hwang SE, Kim SW, Noh DY. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* 2006;**6**: 92.
42. Loo LW, Grove DI, Williams EM, Neal CL, Cousens LA, Schubert EL, Holcomb IN, Massa HF, Glogovac J, Li CI, Malone KE, Daling JR, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res* 2004;**64**: 8541-9.
43. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shtyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 2011;**121**: 2750-67.
44. Silwal-Pandit L, Volland HK, Chin SF, Rueda OM, McKinney S, Osako T, Quigley DA, Kristensen VN, Aparicio S, Borresen-Dale AL, Caldas C, Langerod A. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clin Cancer Res* 2014;**20**: 3569-80.
45. Watkins JA, Irshad S, Grigoriadis A, Tutt AN. Genomic scars as biomarkers of homologous recombination deficiency and drug response in breast and ovarian cancers. *Breast Cancer Res* 2014;**16**: 211.
46. Bettaieb A, Paul C, Plenchette S, Shan J, Chouchane L, Ghiringhelli F. Precision medicine in breast cancer: reality or utopia? *J Transl Med* 2017;**15**: 139.
47. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ, Panel m. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013;**24**: 2206-23.
48. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, Panel m. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011;**22**: 1736-47.
49. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, Davies SR, Snider J, Stijleman IJ, Reed J, Cheang MC, Mardis ER, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res* 2010;**16**: 5222-32.
50. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Bryant J, Costantino JP, Geyer CE, Jr., et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**: 3726-34.
51. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**: 1999-2009.
52. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;**351**: 2817-26.
53. Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, Erlander MG, Dunbier A, Sidhu K, Lopez-Knowles E, Goss PE, Dowsett M. Prediction of late distant recurrence in patients with

oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol* 2013;**14**: 1067-76.

54. Muller BM, Keil E, Lehmann A, Winzer KJ, Richter-Ehrenstein C, Prinzler J, Bangemann N, Reles A, Stadie S, Schoenegg W, Eucker J, Schmidt M, et al. The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. *PLoS One* 2013;**8**: e68252.

55. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;**136**: E359-86.

56. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;**49**: 1374-403.

57. Binder-Foucard F, Bossard N, Delafosse P, Belot A, Woronoff AS, Remontet L, French network of cancer r. Cancer incidence and mortality in France over the 1980-2012 period: solid tumors. *Rev Epidemiol Sante Publique* 2014;**62**: 95-108.

58. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiol Rev* 1993;**15**: 36-47.

59. Clavel-Chapelon F, Gerber M. Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? *Breast Cancer Res Treat* 2002;**72**: 107-15.

60. Albrektsen G, Heuch I, Tretli S, Kvale G. Breast cancer incidence before age 55 in relation to parity and age at first and last births: a prospective study of one million Norwegian women. *Epidemiology* 1994;**5**: 604-11.

61. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer* 2006;**6**: 281-91.

62. Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, Vierkant RA, Maloney SD, Pankratz VS, Hillman DW, Suman VJ, Johnson J, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005;**353**: 229-37.

63. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2006;**15**: 1159-69.

64. Pettersson A, Graff RE, Ursin G, Santos Silva ID, McCormack V, Baglietto L, Vachon C, Bakker MF, Giles GG, Chia KS, Czene K, Eriksson L, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst* 2014;**106**.

65. Lacassagne A. Results of the Treatment of Cancer of the Cervix Uteri. *Br Med J* 1932;**2**: 912-3.

66. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* 1997;**350**: 1047-59.

67. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;**347**: 1713-27.

68. Kumle M, Weiderpass E, Braaten T, Persson I, Adami HO, Lund E. Use of oral contraceptives and breast cancer risk: The Norwegian-Swedish Women's Lifestyle and Health Cohort Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2002;**11**: 1375-81.

69. Claus EB, Stowe M, Carter D. Oral contraceptives and the risk of ductal breast carcinoma in situ. *Breast Cancer Res Treat* 2003;**81**: 129-36.

70. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, Bernstein L, Malone KE, Ursin G, Strom BL, Norman SA, Wingo PA, et al. Oral contraceptives and the risk of breast cancer. *N Engl J Med* 2002;**346**: 2025-32.

71. Cogliano V, Grosse Y, Baan R, Straif K, Secretan B, El Ghissassi F, Cancer WHOIAfRo. Carcinogenicity of combined oestrogen-progestagen contraceptives and menopausal treatment. *Lancet Oncol* 2005;**6**: 552-3.
72. Greenberg ER, Barnes AB, Resseguie L, Barrett JA, Burnside S, Lanza LL, Neff RK, Stevens M, Young RH, Colton T. Breast cancer in mothers given diethylstilbestrol in pregnancy. *N Engl J Med* 1984;**311**: 1393-8.
73. Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, Kaufman R, Herbst AL, Noller KL, Hyer M, Hoover RN. Prenatal diethylstilbestrol exposure and risk of breast cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2006;**15**: 1509-14.
74. UNSCEAR 2000. The United Nations Scientific Committee on the Effects of Atomic Radiation. *Health Phys* 2000;**79**: 314.
75. Ronckers CM, Erdmann CA, Land CE. Radiation and breast cancer: a review of current evidence. *Breast Cancer Res* 2005;**7**: 21-32.
76. Travis LB, Hill DA, Dores GM, Gospodarowicz M, van Leeuwen FE, Holowaty E, Glimelius B, Andersson M, Wiklund T, Lynch CF, Van't Veer MB, Glimelius I, et al. Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. *JAMA* 2003;**290**: 465-75.
77. Inskip PD, Robison LL, Stovall M, Smith SA, Hammond S, Mertens AC, Whitton JA, Diller L, Kenney L, Donaldson SS, Meadows AT, Neglia JP. Radiation dose and breast cancer risk in the childhood cancer survivor study. *J Clin Oncol* 2009;**27**: 3901-7.
78. Guibout C, Adjadj E, Rubino C, Shamsaldin A, Grimaud E, Hawkins M, Mathieu MC, Oberlin O, Zucker JM, Panis X, Lagrange JL, Daly-Schveitzer N, et al. Malignant breast tumors after radiotherapy for a first cancer during childhood. *J Clin Oncol* 2005;**23**: 197-204.
79. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., Coates RJ, Liff JM, Talamini R, Chantarakul N, Koetsawang S, Rachawat D, et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *British journal of cancer* 2002;**87**: 1234-45.
80. Knight JA, Fan J, Malone KE, John EM, Lynch CF, Langballe R, Bernstein L, Shore RE, Brooks JD, Reiner AS, Woods M, Liang X, et al. Alcohol consumption and cigarette smoking in combination: A predictor of contralateral breast cancer risk in the WECARE study. *Int J Cancer* 2017;**141**: 916-24.
81. Schmidt S, Monk JM, Robinson LE, Mourtzakis M. The integrative role of leptin, oestrogen and the insulin family in obesity-associated breast cancer: potential effects of exercise. *Obes Rev* 2015;**16**: 473-87.
82. Renehan AG, Roberts DL, Dive C. Obesity and cancer: pathophysiological and biological mechanisms. *Arch Physiol Biochem* 2008;**114**: 71-83.
83. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D, Million Women Study C. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007;**335**: 1134.
84. Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. *CA Cancer J Clin* 2014;**64**: 186-94.
85. Collaborative Group on Hormonal Factors in Breast C. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 2001;**358**: 1389-99.
86. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990;**250**: 1684-9.
87. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;**266**: 66-71.
88. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. *Nat Rev Cancer* 2004;**4**: 665-76.

89. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* 2011;**12**: 68-78.
90. Kawai H, Li H, Chun P, Avraham S, Avraham HK. Direct interaction between BRCA1 and the estrogen receptor regulates vascular endothelial growth factor (VEGF) transcription and secretion in breast cancer cells. *Oncogene* 2002;**21**: 7730-9.
91. Yarden RI, Pardo-Reoyo S, Sgagias M, Cowan KH, Brody LC. BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. *Nat Genet* 2002;**30**: 285-9.
92. Morris JR, Solomon E. BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Human molecular genetics* 2004;**13**: 807-17.
93. Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, Ogata H, Ohta T. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J Biol Chem* 2001;**276**: 14537-40.
94. Hilbers FS, Vreeswijk MP, van Asperen CJ, Devilee P. The impact of next generation sequencing on the analysis of breast cancer susceptibility: a role for extremely rare genetic variation? *Clin Genet* 2013;**84**: 407-14.
95. Caputo S, Benboudjema L, Sinilnikova O, Rouleau E, Beroud C, Lidereau R, French BGGCC. Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. *Nucleic Acids Res* 2012;**40**: D992-1002.
96. Struwing JP, Abeliovich D, Peretz T, Avishai N, Kaback MM, Collins FS, Brody LC. The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. *Nat Genet* 1995;**11**: 198-200.
97. Neuhausen S, Gilewski T, Norton L, Tran T, McGuire P, Swensen J, Hampel H, Borgen P, Brown K, Skolnick M, Shattuck-Eidens D, Jhanwar S, et al. Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. *Nat Genet* 1996;**13**: 126-8.
98. Tonin PN, Mes-Masson AM, Futreal PA, Morgan K, Mahon M, Foulkes WD, Cole DE, Provencher D, Ghadirian P, Narod SA. Founder BRCA1 and BRCA2 mutations in French Canadian breast and ovarian cancer families. *American journal of human genetics* 1998;**63**: 1341-51.
99. Tonin PN. [The limited spectrum of pathogenic BRCA1 and BRCA2 mutations in the French Canadian breast and breast-ovarian cancer families, a founder population of Quebec, Canada]. *Bull Cancer* 2006;**93**: 841-6.
100. Menkiszak J, Gronwald J, Gorski B, Jakubowska A, Huzarski T, Byrski T, Foszczynska-Kloda M, Haus O, Janiszewska H, Perkowska M, Brozek I, Grzybowska E, et al. Hereditary ovarian cancer in Poland. *Int J Cancer* 2003;**106**: 942-5.
101. Grzybowska E, Zientek H, Jasinska A, Rusin M, Kozlowski P, Sobczak K, Sikorska A, Kwiatkowska E, Gorniak L, Kalinowska E, Utracka-Hutka B, Wloch J, et al. High frequency of recurrent mutations in BRCA1 and BRCA2 genes in Polish families with breast and ovarian cancer. *Human mutation* 2000;**16**: 482-90.
102. Ford D, Easton DF, Peto J. Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. *American journal of human genetics* 1995;**57**: 1457-62.
103. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, 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. *American journal of human genetics* 2003;**72**: 1117-30.
104. Tai YC, Domchek S, Parmigiani G, Chen S. Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst* 2007;**99**: 1811-4.
105. Erkkö H, Xia B, Nikkila J, Schleutker J, Syrjäkoski K, Mannermaa A, Kallioniemi A, Pylkas K, Karppinen SM, Rapakko K, Miron A, Sheng Q, et al. A recurrent mutation in PALB2 in Finnish cancer families. *Nature* 2007;**446**: 316-9.
106. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 2007;**39**: 165-7.

107. Erkkö H, Dowty JG, Nikkila J, Syrjäkoski K, Mannermaa A, Pylkas K, Southey MC, Holli K, Kallioniemi A, Jukkola-Vuorinen A, Kataja V, Kosma VM, et al. Penetrance analysis of the PALB2 c.1592delT founder mutation. *Clin Cancer Res* 2008;**14**: 4667-71.
108. Foulkes WD, Metcalfe K, Sun P, Hanna WM, Lynch HT, Ghadirian P, Tung N, Olopade OI, Weber BL, McLennan J, Olivetto IA, Begin LR, et al. Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influence of age, grade, and histological type. *Clin Cancer Res* 2004;**10**: 2029-34.
109. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leener K, Fostira F, Tomiak E, Neuhausen SL, et al. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* 2014;**371**: 497-506.
110. Malkin D, Li FP, Strong LC, Fraumeni JF, Jr., Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 1990;**250**: 1233-8.
111. Chompret A, Brugieres L, Ronsin M, Gardes M, Dessarps-Freichay F, Abel A, Hua D, Ligt L, Dondon MG, Bressac-de Paillerets B, Frebourg T, Lemerle J, et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *British journal of cancer* 2000;**82**: 1932-7.
112. Tinat J, Bougeard G, Baert-Desurmont S, Vasseur S, Martin C, Bouvignies E, Caron O, Bressac-de Paillerets B, Berthet P, Dugast C, Bonaiti-Pellie C, Stoppa-Lyonnet D, et al. 2009 version of the Chompret criteria for Li Fraumeni syndrome. *J Clin Oncol* 2009;**27**: e108-9; author reply e10.
113. Nelen MR, Kremer H, Konings IB, Schoute F, van Essen AJ, Koch R, Woods CG, Fryns JP, Hamel B, Hoefsloot LH, Peeters EA, Padberg GW. Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. *Eur J Hum Genet* 1999;**7**: 267-73.
114. Bubien V, Bonnet F, Brouste V, Hoppe S, Barouk-Simonet E, David A, Edery P, Bottani A, Layet V, Caron O, Gilbert-Dussardier B, Delnatte C, et al. High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. *J Med Genet* 2013;**50**: 255-63.
115. Giardiello FM, Trimbath JD. Peutz-Jeghers syndrome and management recommendations. *Clin Gastroenterol Hepatol* 2006;**4**: 408-15.
116. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJ, Keller JJ, Westerman AM, Scott RJ, Lim W, Trimbath JD, Giardiello FM, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. *Clin Cancer Res* 2006;**12**: 3209-15.
117. Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, Taite H, Scoular R, Miller A, Reeve AE. E-cadherin germline mutations in familial gastric cancer. *Nature* 1998;**392**: 402-5.
118. Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, Harris LN, Pinheiro HC, Troussard A, Miron P, Tung N, Oliveira C, et al. Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet* 2007;**44**: 726-31.
119. Frank B, Bermejo JL, Hemminki K, Klaes R, Bugert P, Wappenschmidt B, Schmutzler RK, Burwinkel B. Re: Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J Natl Cancer Inst* 2005;**97**: 1012; author reply -3.
120. MacPherson G, Healey CS, Teare MD, Balasubramanian SP, Reed MW, Pharoah PD, Ponder BA, Meuth M, Bhattacharyya NP, Cox A. Association of a common variant of the CASP8 gene with reduced risk of breast cancer. *J Natl Cancer Inst* 2004;**96**: 1866-9.
121. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 2007;**39**: 352-8.
122. Bahcall OG. iCOGS collection provides a collaborative model. Foreword. *Nature genetics* 2013;**45**: 343.
123. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, Edwards SL, Pickett HA, Shen HC, Smart CE, Hillman KM, Mai PL, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* 2013;**45**: 371-84, 84e1-2.

124. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;**47**: 373-80.
125. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;**45**: 353-61, 61e1-2.
126. Antoniou AC, Kartsonaki C, Sinilnikova OM, Soucy P, McGuffog L, Healey S, Lee A, Peterlongo P, Manoukian S, Peissel B, Zaffaroni D, Cattaneo E, et al. Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Human molecular genetics* 2011;**20**: 3304-21.
127. Antoniou AC, Kuchenbaecker KB, Soucy P, Beesley J, Chen X, McGuffog L, Lee A, Barrowdale D, Healey S, Sinilnikova OM, Caligo MA, Loman N, et al. Common variants at 12p11, 12q24, 9p21, 9q31.2 and in ZNF365 are associated with breast cancer risk for BRCA1 and/or BRCA2 mutation carriers. *Breast Cancer Res* 2012;**14**: R33.
128. Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, Versmold B, Engel C, Meindl A, Arnold N, Hofmann W, Sutter C, et al. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *American journal of human genetics* 2008;**82**: 937-48.
129. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009;**41**: 585-90.
130. Lesueur F. [Our trials, sources of transformation]. *Krankenpfl Soins Infirmit* 2014;**107**: 74-5.
131. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, Casey G, Hunter DJ, Sellers TA, Gruber SB, Dunning AM, Michailidou K, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2017;**26**: 126-35.
132. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrulis IL, Terry MB, Daly M, Buys S, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *American journal of human genetics* 2012;**90**: 734-9.
133. Park DJ, Tao K, Le Calvez-Kelm F, Nguyen-Dumont T, Robinot N, Hammet F, Odefrey F, Tsimiklis H, Teo ZL, Thingholm LB, Young EL, Voegelé C, et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov* 2014;**4**: 804-15.
134. Ghoussaini M, Pharoah PD, Easton DF. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? *The American journal of pathology* 2013;**183**: 1038-51.
135. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* 2006;**38**: 873-5.
136. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang SC, Forey N, Feuchtinger C, Gioia L, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *American journal of human genetics* 2009;**85**: 427-46.
137. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, Terry MB, Daly MJ, Buys SM, Southey MC, Andrulis I, John EM, et al. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Res* 2011;**13**: R73.
138. Bernstein JL, Teraoka S, Southey MC, Jenkins MA, Andrulis IL, Knight JA, John EM, Lapinski R, Wolitzer AL, Whittemore AS, West D, Seminara D, et al. Population-based estimates of breast cancer risks associated with ATM gene variants c.7271T>G and c.1066-6T>G (IVS10-6T>G) from the Breast Cancer Family Registry. *Human mutation* 2006;**27**: 1122-8.

139. Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegelé C, Young EL, Robinot N, Forey N, Durand G, Vallee MP, Tao K, Roane TC, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* 2014;**16**: R58.
140. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegelé C, Babikyan D, Durand G, Forey N, McKay-Chopin S, Robinot N, Nguyen-Dumont T, Thomas A, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res* 2011;**13**: R6.
141. Chekmariova EV, Sokolenko AP, Buslov KG, Iyevleva AG, Ulibina YM, Rozanov ME, Mitiushkina NV, Togo AV, Matsko DE, Voskresenskiy DA, Chagunava OL, Devilee P, et al. CHEK2 1100delC mutation is frequent among Russian breast cancer patients. *Breast Cancer Res Treat* 2006;**100**: 99-102.
142. Naslund-Koch C, Nordestgaard BG, Bojesen SE. Increased Risk for Other Cancers in Addition to Breast Cancer for CHEK2*1100delC Heterozygotes Estimated From the Copenhagen General Population Study. *J Clin Oncol* 2016;**34**: 1208-16.
143. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol* 2008;**26**: 542-8.
144. Cybulski C, Wokolorczyk D, Jakubowska A, Huzarski T, Byrski T, Gronwald J, Masojc B, Deebniak T, Gorski B, Blecharz P, Narod SA, Lubinski J. Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. *J Clin Oncol* 2011;**29**: 3747-52.
145. Kilpivaara O, Vahteristo P, Falck J, Syrjakoski K, Eerola H, Easton D, Bartkova J, Lukas J, Heikkila P, Aittomaki K, Holli K, Blomqvist C, et al. CHEK2 variant I157T may be associated with increased breast cancer risk. *Int J Cancer* 2004;**111**: 543-7.
146. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, Hartmann L, Schaal H, Ramser J, Honisch E, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 2010;**42**: 410-4.
147. Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, Bowden G, Kalmyrzaev B, Warren-Perry M, Snape K, Adlard JW, Barwell J, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet* 2011;**43**: 879-82.
148. Lose F, Lovelock P, Chenevix-Trench G, Mann GJ, Pupo GM, Spurdle AB, Kathleen Cuningham Foundation Consortium for Research into Familial Breast C. Variation in the RAD51 gene and familial breast cancer. *Breast Cancer Res* 2006;**8**: R26.
149. Pelttari LM, Kiiski JI, Ranta S, Vilske S, Blomqvist C, Aittomaki K, Nevanlinna H. RAD51, XRCC3, and XRCC2 mutation screening in Finnish breast cancer families. *Springerplus* 2015;**4**: 92.
150. Rapakko K, Heikkinen K, Karppinen SM, Winqvist R. Screening for RAD51 and BRCA2 BRC repeat mutations in breast and ovarian cancer families. *Cancer Lett* 2006;**236**: 142-7.
151. Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, Voegelé C, Vallee MP, Byrnes G, Registry BC, Hopper JL, Southey MC, et al. RAD51 and breast cancer susceptibility: no evidence for rare variant association in the Breast Cancer Family Registry study. *PLoS One* 2012;**7**: e52374.
152. Hilbers FS, Wijnen JT, Hoogerbrugge N, Oosterwijk JC, Collee MJ, Peterlongo P, Radice P, Manoukian S, Feroce I, Capra F, Couch FJ, Wang X, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. *J Med Genet* 2012;**49**: 618-20.
153. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R, Thomas A, Lilyquist J, Feng B, McFarland R, et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol* 2017.
154. Li N, Thompson ER, Rowley SM, McInerney S, Devereux L, Goode D, Investigators L, Wong-Brown MW, Scott RJ, Trainer AH, Goringe KL, James PA, et al. Reevaluation of RINT1 as a breast cancer predisposition gene. *Breast Cancer Res Treat* 2016;**159**: 385-92.

155. Renault AL, Lesueur F, Coulombe Y, Gobeil S, Soucy P, Hamdi Y, Desjardins S, Le Calvez-Kelm F, Vallee M, Voegelé C, Breast Cancer Family R, Hopper JL, et al. ABRAXAS (FAM175A) and Breast Cancer Susceptibility: No Evidence of Association in the Breast Cancer Family Registry. *PLoS One* 2016;**11**: e0156820.
156. Litman R, Peng M, Jin Z, Zhang F, Zhang J, Powell S, Andreassen PR, Cantor SB. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCF. *Cancer Cell* 2005;**8**: 255-65.
157. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 2006;**38**: 1239-41.
158. Easton DF, Lesueur F, Decker B, Michailidou K, Li J, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK, Wang Q, Dennis J, et al. No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk: implications for gene panel testing. *J Med Genet* 2016;**53**: 298-309.
159. Watson P, Lynch HT. Cancer risk in mismatch repair gene mutation carriers. *Fam Cancer* 2001;**1**: 57-60.
160. Savitsky K, Sfez S, Tagle DA, Ziv Y, Sartiell A, Collins FS, Shiloh Y, Rotman G. The complete sequence of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Human molecular genetics* 1995;**4**: 2025-32.
161. Bensimon A, Schmidt A, Ziv Y, Elkon R, Wang SY, Chen DJ, Aebersold R, Shiloh Y. ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci Signal* 2010;**3**: rs3.
162. Bakkenist CJ, Kastan MB. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* 2003;**421**: 499-506.
163. Savic V, Yin B, Maas NL, Bredemeyer AL, Carpenter AC, Helmink BA, Yang-Iott KS, Sleckman BP, Bassing CH. Formation of dynamic gamma-H2AX domains along broken DNA strands is distinctly regulated by ATM and MDC1 and dependent upon H2AX densities in chromatin. *Molecular cell* 2009;**34**: 298-310.
164. Caestecker KW, Van de Walle GR. The role of BRCA1 in DNA double-strand repair: past and present. *Exp Cell Res* 2013;**319**: 575-87.
165. Lavin MF, Kozlov S. ATM activation and DNA damage response. *Cell Cycle* 2007;**6**: 931-42.
166. Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* 2013;**14**: 197-210.
167. Swift M, Morrell D, Cromartie E, Chamberlin AR, Skolnick MH, Bishop DT. The incidence and gene frequency of ataxia-telangiectasia in the United States. *American journal of human genetics* 1986;**39**: 573-83.
168. Gatti RA. Ataxia-telangiectasia. *Dermatol Clin* 1995;**13**: 1-6.
169. Boder E. Ataxia-telangiectasia: an overview. *Kroc Found Ser* 1985;**19**: 1-63.
170. Greenberger S, Berkun Y, Ben-Zeev B, Levi YB, Barzilai A, Nissenkorn A. Dermatologic manifestations of ataxia-telangiectasia syndrome. *J Am Acad Dermatol* 2013;**68**: 932-6.
171. Morrell D, Cromartie E, Swift M. Mortality and cancer incidence in 263 patients with ataxia-telangiectasia. *J Natl Cancer Inst* 1986;**77**: 89-92.
172. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**: 1073-81.
173. Li A, Swift M. Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients. *Am J Med Genet* 2000;**92**: 170-7.
174. Barone G, Groom A, Reiman A, Srinivasan V, Byrd PJ, Taylor AM. Modeling ATM mutant proteins from missense changes confirms retained kinase activity. *Human mutation* 2009;**30**: 1222-30.
175. Taylor AM, Edwards MJ. Malignancy, DNA damage and chromosomal aberrations in ataxia telangiectasia. *IARC Sci Publ* 1982: 119-26.

176. Swift M, Reitnauer PJ, Morrell D, Chase CL. Breast and other cancers in families with ataxia-telangiectasia. *N Engl J Med* 1987;**316**: 1289-94.
177. Athma P, Rappaport R, Swift M. Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. *Cancer Genet Cytogenet* 1996;**92**: 130-4.
178. Borresen AL, Andersen TI, Tretli S, Heiberg A, Moller P. Breast cancer and other cancers in Norwegian families with ataxia-telangiectasia. *Genes Chromosomes Cancer* 1990;**2**: 339-40.
179. Janin N, Andrieu N, Ossian K, Lauge A, Croquette MF, Griscelli C, Debre M, Bressac-de-Paillerets B, Aurias A, Stoppa-Lyonnet D. Breast cancer risk in ataxia telangiectasia (AT) heterozygotes: haplotype study in French AT families. *British journal of cancer* 1999;**80**: 1042-5.
180. Morrell D, Chase CL, Swift M. Cancers in 44 families with ataxia-telangiectasia. *Cancer Genet Cytogenet* 1990;**50**: 119-23.
181. Geoffroy-Perez B, Janin N, Ossian K, Lauge A, Croquette MF, Griscelli C, Debre M, Bressac-de-Paillerets B, Aurias A, Stoppa-Lyonnet D, Andrieu N. Cancer risk in heterozygotes for ataxia-telangiectasia. *Int J Cancer* 2001;**93**: 288-93.
182. Cavaciuti E, Lauge A, Janin N, Ossian K, Hall J, Stoppa-Lyonnet D, Andrieu N. Cancer risk according to type and location of ATM mutation in ataxia-telangiectasia families. *Genes Chromosomes Cancer* 2005;**42**: 1-9.
183. Broeks A, Urbanus JH, Floore AN, Dahler EC, Klijn JG, Rutgers EJ, Devilee P, Russell NS, van Leeuwen FE, van 't Veer LJ. ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. *American journal of human genetics* 2000;**66**: 494-500.
184. Chen J, Birkholtz GG, Lindblom P, Rubio C, Lindblom A. The role of ataxia-telangiectasia heterozygotes in familial breast cancer. *Cancer Res* 1998;**58**: 1376-9.
185. FitzGerald MG, Bean JM, Hegde SR, Unsal H, MacDonald DJ, Harkin DP, Finkelstein DM, Isselbacher KJ, Haber DA. Heterozygous ATM mutations do not contribute to early onset of breast cancer. *Nat Genet* 1997;**15**: 307-10.
186. Andrieu N, Cavaciuti E, Lauge A, Ossian K, Janin N, Hall J, Stoppa-Lyonnet D. Ataxia-telangiectasia genes and breast cancer risk in a French family study. *The Journal of dairy research* 2005;**72 Spec No**: 73-80.
187. Ng PC. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* 2003;**31**: 3812-4.
188. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nature methods* 2010;**7**: 248-9.
189. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* 2010;**20**: 110-21.
190. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB, Group IUGVW. In silico analysis of missense substitutions using sequence-alignment based methods. *Human mutation* 2008;**29**: 1327-36.
191. Seal S, Barfoot R, Jayatilake H, Smith P, Renwick A, Bascombe L, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N, et al. Evaluation of Fanconi Anemia genes in familial breast cancer predisposition. *Cancer Res* 2003;**63**: 8596-9.
192. Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, Sampson JR, Cheadle JP. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C-->T:A mutations. *Human molecular genetics* 2002;**11**: 2961-7.
193. Wasielewski M, Out AA, Vermeulen J, Nielsen M, van den Ouweland A, Tops CM, Wijnen JT, Vasen HF, Weiss MM, Klijn JG, Devilee P, Hes FJ, et al. Increased MUTYH mutation frequency among Dutch families with breast cancer and colorectal cancer. *Breast Cancer Res Treat* 2010;**124**: 635-41.
194. Win AK, Reece JC, Dowty JG, Buchanan DD, Clendenning M, Rosty C, Southey MC, Young JP, Cleary SP, Kim H, Cotterchio M, Macrae FA, et al. Risk of extracolonic cancers for people with biallelic and monoallelic mutations in MUTYH. *Int J Cancer* 2016;**139**: 1557-63.

195. Out AA, Wasielewski M, Huijts PE, van Minderhout IJ, Houwing-Duistermaat JJ, Tops CM, Nielsen M, Seynaeve C, Wijnen JT, Breuning MH, van Asperen CJ, Schutte M, et al. MUTYH gene variants and breast cancer in a Dutch case-control study. *Breast Cancer Res Treat* 2012;**134**: 219-27.
196. Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, Igari T, Koike M, Chiba M, Mori T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* 1997;**17**: 271-2.
197. Wijnen J, de Leeuw W, Vasen H, van der Klift H, Moller P, Stormorken A, Meijers-Heijboer H, Lindhout D, Menko F, Vossen S, Moslein G, Tops C, et al. Familial endometrial cancer in female carriers of MSH6 germline mutations. *Nat Genet* 1999;**23**: 142-4.
198. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, et al. Clues to the pathogenesis of familial colorectal cancer. *Science* 1993;**260**: 812-6.
199. Murphy KM, Zhang S, Geiger T, Hafez MJ, Bacher J, Berg KD, Eshleman JR. Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers. *J Mol Diagn* 2006;**8**: 305-11.
200. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 1988;**85**: 6622-6.
201. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005;**19**: 2100-10.
202. Palm W, de Lange T. How shelterin protects mammalian telomeres. *Annu Rev Genet* 2008;**42**: 301-34.
203. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature* 1990;**345**: 458-60.
204. Blackburn EH. Structure and function of telomeres. *Nature* 1991;**350**: 569-73.
205. Cervantes RB, Lundblad V. Mechanisms of chromosome-end protection. *Curr Opin Cell Biol* 2002;**14**: 351-6.
206. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. Mammalian telomeres end in a large duplex loop. *Cell* 1999;**97**: 503-14.
207. Doksani Y, Wu JY, de Lange T, Zhuang X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* 2013;**155**: 345-56.
208. Ancelin K, Brunori M, Bauwens S, Koering CE, Brun C, Ricoul M, Pommier JP, Sabatier L, Gilson E. Targeting assay to study the cis functions of human telomeric proteins: evidence for inhibition of telomerase by TRF1 and for activation of telomere degradation by TRF2. *Mol Cell Biol* 2002;**22**: 3474-87.
209. van Steensel B, de Lange T. Control of telomere length by the human telomeric protein TRF1. *Nature* 1997;**385**: 740-3.
210. Stansel RM, de Lange T, Griffith JD. T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *EMBO J* 2001;**20**: 5532-40.
211. Zhu XD, Kuster B, Mann M, Petrini JH, de Lange T. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nat Genet* 2000;**25**: 347-52.
212. Li B, de Lange T. Rap1 affects the length and heterogeneity of human telomeres. *Mol Biol Cell* 2003;**14**: 5060-8.
213. Li B, Oestreich S, de Lange T. Identification of human Rap1: implications for telomere evolution. *Cell* 2000;**101**: 471-83.
214. Loayza D, Parsons H, Donigian J, Hoke K, de Lange T. DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *J Biol Chem* 2004;**279**: 13241-8.
215. O'Connor MS, Safari A, Xin H, Liu D, Songyang Z. A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proceedings of the National Academy of Sciences of the United States of America* 2006;**103**: 11874-9.

216. Houghtaling BR, Cuttonaro L, Chang W, Smith S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr Biol* 2004;**14**: 1621-31.
217. Ye JZ, Donigian JR, van Overbeek M, Loayza D, Luo Y, Krutchinsky AN, Chait BT, de Lange T. TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J Biol Chem* 2004;**279**: 47264-71.
218. Ye JZ, Hockemeyer D, Krutchinsky AN, Loayza D, Hooper SM, Chait BT, de Lange T. POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev* 2004;**18**: 1649-54.
219. Loayza D, De Lange T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* 2003;**423**: 1013-8.
220. Xin H, Liu D, Wan M, Safari A, Kim H, Sun W, O'Connor MS, Songyang Z. TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* 2007;**445**: 559-62.
221. Liu D, Safari A, O'Connor MS, Chan DW, Laegerler A, Qin J, Songyang Z. PTOP interacts with POT1 and regulates its localization to telomeres. *Nat Cell Biol* 2004;**6**: 673-80.
222. de Lange T. How telomeres solve the end-protection problem. *Science* 2009;**326**: 948-52.
223. Karlseder J, Hoke K, Mirzoeva OK, Bakkenist C, Kastan MB, Petrini JH, de Lange T. The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. *PLoS Biol* 2004;**2**: E240.
224. Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* 2009;**138**: 90-103.
225. Schmutz I, de Lange T. Shelterin. *Curr Biol* 2016;**26**: R397-9.
226. Wu Y, Xiao S, Zhu XD. MRE11-RAD50-NBS1 and ATM function as co-mediators of TRF1 in telomere length control. *Nat Struct Mol Biol* 2007;**14**: 832-40.
227. Greider CW, Blackburn EH. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* 1989;**337**: 331-7.
228. Maciejowski J, de Lange T. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol* 2017;**18**: 175-86.
229. Weinrich SL, Pruzan R, Ma L, Ouellette M, Tesmer VM, Holt SE, Bodnar AG, Lichtsteiner S, Kim NW, Trager JB, Taylor RD, Carlos R, et al. Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTERT. *Nat Genet* 1997;**17**: 498-502.
230. Calado RT, Young NS. Telomere diseases. *N Engl J Med* 2009;**361**: 2353-65.
231. Holohan B, Wright WE, Shay JW. Cell biology of disease: Telomeropathies: an emerging spectrum disorder. *J Cell Biol* 2014;**205**: 289-99.
232. Mason PJ, Bessler M. The genetics of dyskeratosis congenita. *Cancer Genet* 2011;**204**: 635-45.
233. Vulliamy TJ, Knight SW, Mason PJ, Dokal I. Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita. *Blood Cells Mol Dis* 2001;**27**: 353-7.
234. Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 2001;**413**: 432-5.
235. Wong JM, Collins K. Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. *Genes Dev* 2006;**20**: 2848-58.
236. Wong JM, Kyasa MJ, Hutchins L, Collins K. Telomerase RNA deficiency in peripheral blood mononuclear cells in X-linked dyskeratosis congenita. *Hum Genet* 2004;**115**: 448-55.
237. Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, Greider CW. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proceedings of the National Academy of Sciences of the United States of America* 2005;**102**: 15960-4.

238. Sharma A, Myers K, Ye Z, D'Orazio J. Dyskeratosis congenita caused by a novel TERT point mutation in siblings with pancytopenia and exudative retinopathy. *Pediatr Blood Cancer* 2014;**61**: 2302-4.
239. Stockklauser C, Raffel S, Klermund J, Bandapalli OR, Beier F, Brummendorf TH, Burger F, Sauer SW, Hoffmann GF, Lorenz H, Tagliaferri L, Nowak D, et al. A novel autosomal recessive TERT T1129P mutation in a dyskeratosis congenita family leads to cellular senescence and loss of CD34+ hematopoietic stem cells not reversible by mTOR-inhibition. *Aging (Albany NY)* 2015;**7**: 911-27.
240. Savage SA, Giri N, Baerlocher GM, Orr N, Lansdorp PM, Alter BP. TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. *American journal of human genetics* 2008;**82**: 501-9.
241. Sasa GS, Ribes-Zamora A, Nelson ND, Bertuch AA. Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood. *Clin Genet* 2012;**81**: 470-8.
242. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA, 3rd, Lansdorp PM, Greider CW, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 2007;**356**: 1317-26.
243. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, Rosenblatt RL, Shay JW, Garcia CK. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proceedings of the National Academy of Sciences of the United States of America* 2007;**104**: 7552-7.
244. Petrovski S, Todd JL, Durham MT, Wang Q, Chien JW, Kelly FL, Frankel C, Mebane CM, Ren Z, Bridgers J, Urban TJ, Malone CD, et al. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2017.
245. Ishikawa N, Nakamura K, Izumiyama-Shimomura N, Aida J, Ishii A, Goto M, Ishikawa Y, Asaka R, Matsuura M, Hatamochi A, Kuroiwa M, Takubo K. Accelerated in vivo epidermal telomere loss in Werner syndrome. *Aging (Albany NY)* 2011;**3**: 417-29.
246. Tahara H, Tokutake Y, Maeda S, Kataoka H, Watanabe T, Satoh M, Matsumoto T, Sugawara M, Ide T, Goto M, Furuichi Y, Sugimoto M. Abnormal telomere dynamics of B-lymphoblastoid cell strains from Werner's syndrome patients transformed by Epstein-Barr virus. *Oncogene* 1997;**15**: 1911-20.
247. Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A, Pak E, et al. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* 2003;**423**: 293-8.
248. d'Adda di Fagagna F, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, Von Zglinicki T, Saretzki G, Carter NP, Jackson SP. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 2003;**426**: 194-8.
249. Artandi SE, Chang S, Lee SL, Alson S, Gottlieb GJ, Chin L, DePinho RA. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* 2000;**406**: 641-5.
250. Davoli T, de Lange T. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer Cell* 2012;**21**: 765-76.
251. Davoli T, Denchi EL, de Lange T. Persistent telomere damage induces bypass of mitosis and tetraploidy. *Cell* 2010;**141**: 81-93.
252. Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* 2015;**163**: 1641-54.
253. Riboni R, Casati A, Nardo T, Zaccaro E, Ferretti L, Nuzzo F, Mondello C. Telomeric fusions in cultured human fibroblasts as a source of genomic instability. *Cancer Genet Cytogenet* 1997;**95**: 130-6.
254. Roger L, Jones RE, Heppel NH, Williams GT, Sampson JR, Baird DM. Extensive telomere erosion in the initiation of colorectal adenomas and its association with chromosomal instability. *J Natl Cancer Inst* 2013;**105**: 1202-11.
255. Shay JW, Reddel RR, Wright WE. Cancer. Cancer and telomeres--an ALternative to telomerase. *Science* 2012;**336**: 1388-90.

256. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K, Schadendorf D, Kumar R. TERT promoter mutations in familial and sporadic melanoma. *Science* 2013;**339**: 959-61.
257. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science* 2013;**339**: 957-9.
258. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA, Jr., Friedman AH, Friedman H, Gallia GL, Giovanella BC, Grollman AP, He TC, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proceedings of the National Academy of Sciences of the United States of America* 2013;**110**: 6021-6.
259. Quaas A, Oldopp T, Tharun L, Klingensfeld C, Krech T, Sauter G, Grob TJ. Frequency of TERT promoter mutations in primary tumors of the liver. *Virchows Arch* 2014;**465**: 673-7.
260. Remke M, Ramaswamy V, Peacock J, Shih DJ, Koelsche C, Northcott PA, Hill N, Cavalli FM, Kool M, Wang X, Mack SC, Barszczyk M, et al. TERT promoter mutations are highly recurrent in SHH subgroup medulloblastoma. *Acta Neuropathol* 2013;**126**: 917-29.
261. Borah S, Xi L, Zaugh AJ, Powell NM, Dancik GM, Cohen SB, Costello JC, Theodorescu D, Cech TR. Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. *Science* 2015;**347**: 1006-10.
262. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, Tiffen JC, Petljak M, Palmer JM, Symmons J, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet* 2014;**46**: 478-81.
263. Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fagnoli MC, Ghiorzo P, Bressac-de Paillerets B, Nagore E, Avril MF, Caporaso NE, McMaster ML, et al. Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat Genet* 2014;**46**: 482-6.
264. Machiela MJ, Lan Q, Slager SL, Vermeulen RC, Teras LR, Camp NJ, Cerhan JR, Spinelli JJ, Wang SS, Nieters A, Vijai J, Yeager M, et al. Genetically predicted longer telomere length is associated with increased risk of B-cell lymphoma subtypes. *Human molecular genetics* 2016;**25**: 1663-76.
265. Ojha J, Codd V, Nelson CP, Samani NJ, Smirnov IV, Madsen NR, Hansen HM, de Smith AJ, Bracci PM, Wiencke JK, Wrensch MR, Wiemels JL, et al. Genetic Variation Associated with Longer Telomere Length Increases Risk of Chronic Lymphocytic Leukemia. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2016;**25**: 1043-9.
266. Machiela MJ, Hsiung CA, Shu XO, Seow WJ, Wang Z, Matsuo K, Hong YC, Seow A, Wu C, Hosgood HD, 3rd, Chen K, Wang JC, et al. Genetic variants associated with longer telomere length are associated with increased lung cancer risk among never-smoking women in Asia: a report from the female lung cancer consortium in Asia. *Int J Cancer* 2015;**137**: 311-9.
267. Kimura M, Aviv A. Measurement of telomere DNA content by dot blot analysis. *Nucleic Acids Res* 2011;**39**: e84.
268. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, Harley CB, Aviv A. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc* 2010;**5**: 1596-607.
269. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 2002;**30**: e47.
270. Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009;**37**: e21.
271. Ferlicot S, Youssef N, Feneux D, Delhommeau F, Paradis V, Bedossa P. Measurement of telomere length on tissue sections using quantitative fluorescence in situ hybridization (Q-FISH). *J Pathol* 2003;**200**: 661-6.
272. Ourliac-Garnier I, Londono-Vallejo A. Telomere length analysis by quantitative fluorescent in situ hybridization (Q-FISH). *Methods Mol Biol* 2011;**735**: 21-31.
273. Savage SA, Chanock SJ, Lissowska J, Brinton LA, Richesson D, Peplonska B, Bardin-Mikolajczak A, Zatonski W, Szeszenia-Dabrowska N, Garcia-Closas M. Genetic variation in five genes important in telomere biology and risk for breast cancer. *British journal of cancer* 2007;**97**: 832-6.

274. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I, Broer L, Nyholt DR, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013;**45**: 422-7, 7e1-2.
275. Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, Michailidou K, Tyrer JP, Benlloch S, Brown J, Audley T, Luben R, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Human molecular genetics* 2013;**22**: 5056-64.
276. Ghossaini M, French JD, Michailidou K, Nord S, Beesley J, Canisus S, Hillman KM, Kaufmann S, Sivakumaran H, Moradi Marjaneh M, Lee JS, Dennis J, et al. Evidence that the 5p12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGF10 and MRPS30 Regulation. *American journal of human genetics* 2016;**99**: 903-11.
277. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, Strobbe LJ, Swinkels DW, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008;**40**: 703-6.
278. Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS, Apicella C, Smith LD, Hammet F, Southey MC, Van 't Veer LJ, de Groot R, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Human molecular genetics* 2011;**20**: 3289-303.
279. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI, Milne RL, Ribas G, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS genetics* 2008;**4**: e1000054.
280. Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswold C, Fasching PA, Miron P, Eccles D, Carpenter JE, Godwin AK, Ambrosone C, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res* 2011;**71**: 6240-9.
281. Sawyer E, Roylance R, Petridis C, Brook MN, Nowinski S, Papouli E, Fletcher O, Pinder S, Hanby A, Kohut K, Gorman P, Caneppele M, et al. Genetic predisposition to in situ and invasive lobular carcinoma of the breast. *PLoS genetics* 2014;**10**: e1004285.
282. Melin BS, Nordfjall K, Andersson U, Roos G. hTERT cancer risk genotypes are associated with telomere length. *Genetic epidemiology* 2012;**36**: 368-72.
283. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, et al. Lung cancer susceptibility locus at 5p15.33. *Nat Genet* 2008;**40**: 1404-6.
284. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre JY, Hoang-Xuan K, El Hallani S, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* 2009;**41**: 899-904.
285. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 2009;**41**: 579-84.
286. Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC, Apicella C, Hammet F, Schmidt MK, Broeks A, Tollenaar RA, Van't Veer LJ, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Human molecular genetics* 2011;**20**: 4693-706.
287. Antoniou AC, Beesley J, McGuffog L, Sinilnikova OM, Healey S, Neuhausen SL, Ding YC, Rebbeck TR, Weitzel JN, Lynch HT, Isaacs C, Ganz PA, et al. Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. *Cancer Res* 2010;**70**: 9742-54.
288. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, et al. Common variants on

chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007;**39**: 865-9.

289. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009;**41**: 324-8.

290. Mulligan AM, Couch FJ, Barrowdale D, Domchek SM, Eccles D, Nevanlinna H, Ramus SJ, Robson M, Sherman M, Spurdle AB, Wappenschmidt B, Lee A, et al. Common breast cancer susceptibility alleles are associated with tumour subtypes in BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2. *Breast Cancer Res* 2011;**13**: R110.

291. Campa D, Barrdahl M, Gaudet MM, Black A, Chanock SJ, Diver WR, Gapstur SM, Haiman C, Hankinson S, Hazra A, Henderson B, Hoover RN, et al. Genetic risk variants associated with in situ breast cancer. *Breast Cancer Res* 2015;**17**: 82.

292. Silvestri V, Rizzolo P, Scarno M, Chillemi G, Navazio AS, Valentini V, Zelli V, Zanna I, Saieva C, Masala G, Bianchi S, Manoukian S, et al. Novel and known genetic variants for male breast cancer risk at 8q24.21, 9p21.3, 11q13.3 and 14q24.1: results from a multicenter study in Italy. *Eur J Cancer* 2015;**51**: 2289-95.

293. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghousaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010;**42**: 504-7.

294. Wyszynski A, Hong CC, Lam K, Michailidou K, Lytle C, Yao S, Zhang Y, Bolla MK, Wang Q, Dennis J, Hopper JL, Southey MC, et al. An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. *Human molecular genetics* 2016;**25**: 3863-76.

295. Antoniou AC, Sinilnikova OM, McGuffog L, Healey S, Nevanlinna H, Heikkinen T, Simard J, Spurdle AB, Beesley J, Chen X, Kathleen Cuninghame Foundation Consortium for Research into Familial Breast C, Neuhausen SL, et al. Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Human molecular genetics* 2009;**18**: 4442-56.

296. Fletcher O, Johnson N, Gibson L, Coupland B, Fraser A, Leonard A, dos Santos Silva I, Ashworth A, Houlston R, Peto J. Association of genetic variants at 8q24 with breast cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2008;**17**: 702-5.

297. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;**447**: 1087-93.

298. Orr N, Cooke R, Jones M, Fletcher O, Dudbridge F, Chilcott-Burns S, Tomczyk K, Broderick P, Houlston R, Ashworth A, Swerdlow A. Genetic variants at chromosomes 2q35, 5p12, 6q25.1, 10q26.13, and 16q12.1 influence the risk of breast cancer in men. *PLoS genetics* 2011;**7**: e1002290.

299. Armes JE, Trute L, White D, Southey MC, Hammet F, Tesoriero A, Hutchins AM, Dite GS, McCredie MR, Giles GG, Hopper JL, Venter DJ. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. *Cancer Res* 1999;**59**: 2011-7.

300. Eisinger F, Jacquemier J, Charpin C, Stoppa-Lyonnet D, Bressac-de Paillerets B, Peyrat JP, Longy M, Guinebretiere JM, Sauvan R, Noguchi T, Birnbaum D, Sobol H. Mutations at BRCA1: the medullary breast carcinoma revisited. *Cancer Res* 1998;**58**: 1588-92.

301. Iau PT, Marafie M, Ali A, Sng JH, Macmillan RD, Pinder S, Denley HE, Ellis IO, Wenzyc P, Scott N, Cross G, Blamey RW. Are medullary breast cancers an indication for BRCA1 mutation screening? A mutation analysis of 42 cases of medullary breast cancer. *Breast Cancer Res Treat* 2004;**85**: 81-8.

302. Ridolfi RL, Rosen PP, Port A, Kinne D, Mike V. Medullary carcinoma of the breast: a clinicopathologic study with 10 year follow-up. *Cancer* 1977;**40**: 1365-85.

303. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium. *Lancet* 1997;**349**: 1505-10.
304. Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, Farid LM, Venter D, Antoniou A, Storfer-Isser A, Smyth E, Steel CM, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst* 1998;**90**: 1138-45.
305. Tung N, Wang Y, Collins LC, Kaplan J, Li H, Gelman R, Comander AH, Gallagher B, Fetten K, Krag K, Stoeckert KA, Legare RD, et al. Estrogen receptor positive breast cancers in BRCA1 mutation carriers: clinical risk factors and pathologic features. *Breast Cancer Res* 2010;**12**: R12.
306. Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, Easton DF. 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;**20**: 2310-8.
307. Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, Hortobagyi GN, Arun BK. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J Clin Oncol* 2008;**26**: 4282-8.
308. Yadav BS, Chanana P, Jhamb S. Biomarkers in triple negative breast cancer: A review. *World J Clin Oncol* 2015;**6**: 252-63.
309. Mavaddat N, Barrowdale D, Andrulis IL, Domchek SM, Eccles D, Nevanlinna H, Ramus SJ, Spurdle A, Robson M, Sherman M, Mulligan AM, Couch FJ, et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2012;**21**: 134-47.
310. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* 1971;**68**: 820-3.
311. Neuhausen SL, Marshall CJ. Loss of heterozygosity in familial tumors from three BRCA1-linked kindreds. *Cancer Res* 1994;**54**: 6069-72.
312. Lips EH, Debipersad RD, Scheerman CE, Mulder L, Sonke GS, van der Kolk LE, Wesseling J, Hogervorst FB, Nederlof PM. BRCA1-Mutated Estrogen Receptor-Positive Breast Cancer Shows BRCAness, Suggesting Sensitivity to Drugs Targeting Homologous Recombination Deficiency. *Clin Cancer Res* 2017;**23**: 1236-41.
313. Osorio A, de la Hoya M, Rodriguez-Lopez R, Martinez-Ramirez A, Cazorla A, Granizo JJ, Esteller M, Rivas C, Caldes T, Benitez J. Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. *Int J Cancer* 2002;**99**: 305-9.
314. Rzepecka IK, Szafron L, Stys A, Bujko M, Plisiecka-Halasa J, Madry R, Osuch B, Markowska J, Bidzinski M, Kupryjanczyk J. High frequency of allelic loss at the BRCA1 locus in ovarian cancers: clinicopathologic and molecular associations. *Cancer Genet* 2012;**205**: 94-100.
315. Bane AL, Beck JC, Bleiweiss I, Buys SS, Catalano E, Daly MB, Giles G, Godwin AK, Hibshoosh H, Hopper JL, John EM, Layfield L, 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;**31**: 121-8.
316. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, Bishop T, Benitez J, Rivas C, Bignon YJ, Chang-Claude J, Hamann U, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clin Cancer Res* 2005;**11**: 5175-80.
317. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene* 2006;**25**: 5846-53.
318. Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, Karhu R, Tanner M, Isola J, Barkardottir RB, Borg A, Kallioniemi OP. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res* 1997;**57**: 1222-7.

319. Gonzalez-Angulo AM, Timms KM, Liu S, Chen H, Litton JK, Potter J, Lanchbury JS, Stemke-Hale K, Hennessy BT, Arun BK, Hortobagyi GN, Do KA, et al. Incidence and outcome of BRCA mutations in unselected patients with triple receptor-negative breast cancer. *Clin Cancer Res* 2011;**17**: 1082-9.
320. Popova T, Manie E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, Delattre O, Sigal-Zafrani B, Bollet M, Longy M, Houdayer C, Sastre-Garau X, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* 2012;**72**: 5454-62.
321. Popova T, Manie E, Stoppa-Lyonnet D, Rigai G, Barillot E, Stern MH. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome biology* 2009;**10**: R128.
322. Tan DS, Marchio C, Reis-Filho JS. Hereditary breast cancer: from molecular pathology to tailored therapies. *J Clin Pathol* 2008;**61**: 1073-82.
323. Vaca-Paniagua F, Alvarez-Gomez RM, Maldonado-Martinez HA, Perez-Plasencia C, Fragoso-Ontiveros V, Laso-Gonsebatt F, Herrera LA, Cantu D, Bargallo-Rocha E, Mohar A, Durand G, Forey N, et al. Revealing the Molecular Portrait of Triple Negative Breast Tumors in an Understudied Population through Omics Analysis of Formalin-Fixed and Paraffin-Embedded Tissues. *PLoS One* 2015;**10**: e0126762.
324. Palacios J, Honrado E, Osorio A, Cazorla A, Sarrio D, Barroso A, Rodriguez S, Cigudosa JC, Diez O, Alonso C, Lerma E, Dopazo J, et al. Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. *Breast Cancer Res Treat* 2005;**90**: 5-14.
325. Collins N, McManus R, Wooster R, Mangion J, Seal S, Lakhani SR, Ormiston W, Daly PA, Ford D, Easton DF, et al. Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12-13. *Oncogene* 1995;**10**: 1673-5.
326. van Beers EH, van Welsem T, Wessels LF, Li Y, Oldenburg RA, Devilee P, Cornelisse CJ, Verhoef S, Hogervorst FB, van't Veer LJ, Nederlof PM. Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations. *Cancer Res* 2005;**65**: 822-7.
327. Stefansson OA, Jonasson JG, Johannsson OT, Olafsdottir K, Steinarsdottir M, Valgeirsdottir S, Eyfjord JE. Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Res* 2009;**11**: R47.
328. Pecuchet N, Popova T, Manie E, Lucchesi C, Battistella A, Vincent-Salomon A, Caux-Moncoutier V, Bollet M, Sigal-Zafrani B, Sastre-Garau X, Stoppa-Lyonnet D, Stern MH. Loss of heterozygosity at 13q13 and 14q32 predicts BRCA2 inactivation in luminal breast carcinomas. *Int J Cancer* 2013;**133**: 2834-42.
329. Rouault A, Banneau G, Macgrogan G, Jones N, Elarouci N, Barouk-Simonet E, Venat L, Coupier I, Letouze E, de Reynies A, Bonnet F, Iggo R, et al. Deletion of chromosomes 13q and 14q is a common feature of tumors with BRCA2 mutations. *PLoS One* 2012;**7**: e52079.
330. Jonsdottir AB, Stefansson OA, Bjornsson J, Jonasson JG, Ogmundsdottir HM, Eyfjord JE. Tetraploidy in BRCA2 breast tumours. *Eur J Cancer* 2012;**48**: 305-10.
331. Schoolmeester JK, Moyer AM, Goodenberger ML, Keeney GL, Carter JM, Bakkum-Gamez JN. Pathologic Findings in Breast, Fallopian Tube and Ovary Specimens in non-BRCA Hereditary Breast and/or Ovarian Cancer Syndromes: A Study of 18 Patients with Deleterious Germline Mutations in RAD51C, BARD1, BRIP1, PALB2, MUTYH or CHEK2. *Hum Pathol* 2017.
332. Teo ZL, Provenzano E, Dite GS, Park DJ, Apicella C, Sawyer SD, James PA, Mitchell G, Trainer AH, Lindeman GJ, Shackleton K, Cicciarelli L, et al. Tumour morphology predicts PALB2 germline mutation status. *British journal of cancer* 2013;**109**: 154-63.
333. Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, Swiec M, Byrski T, Debniak T, Gorski B, Sopik V, Akbari MR, et al. Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. *Lancet Oncol* 2015;**16**: 638-44.

334. Hartley T, Cavallone L, Sabbaghian N, Silva-Smith R, Hamel N, Aleynikova O, Smith E, Hastings V, Pinto P, Tischkowitz M, Tomiak E, Foulkes WD. Mutation analysis of PALB2 in BRCA1 and BRCA2-negative breast and/or ovarian cancer families from Eastern Ontario, Canada. *Hered Cancer Clin Pract* 2014;**12**: 19.
335. Domagala P, Wokolorczyk D, Cybulski C, Huzarski T, Lubinski J, Domagala W. Different CHEK2 germline mutations are associated with distinct immunophenotypic molecular subtypes of breast cancer. *Breast Cancer Res Treat* 2012;**132**: 937-45.
336. Huzarski T, Cybulski C, Domagala W, Gronwald J, Byrski T, Szwiec M, Woyke S, Narod SA, Lubinski J. Pathology of breast cancer in women with constitutional CHEK2 mutations. *Breast Cancer Res Treat* 2005;**90**: 187-9.
337. Suspitsin EN, Yanus GA, Sokolenko AP, Yatsuk OS, Zaitseva OA, Bessonov AA, Ivantsov AO, Heinstejn VA, Klimashevskiy VF, Togo AV, Imyanitov EN. Development of breast tumors in CHEK2, NBN/NBS1 and BLM mutation carriers does not commonly involve somatic inactivation of the wild-type allele. *Med Oncol* 2014;**31**: 828.
338. Massink MP, Kooi IE, Martens JW, Waisfisz Q, Meijers-Heijboer H. Genomic profiling of CHEK2*1100delC-mutated breast carcinomas. *BMC Cancer* 2015;**15**: 877.
339. Honrado E, Osorio A, Palacios J, Benitez J. Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations. *Oncogene* 2006;**25**: 5837-45.
340. Gevensleben H, Bossung V, Meindl A, Wappenschmidt B, de Gregorio N, Osorio A, Romero A, Buettner R, Markiefka B, Schmutzler RK. Pathological features of breast and ovarian cancers in RAD51C germline mutation carriers. *Virchows Arch* 2014;**465**: 365-9.
341. Schnurbein G, Hauke J, Wappenschmidt B, Weber-Lassalle N, Engert S, Hellebrand H, Garbes L, Becker A, Neidhardt G, Rhiem K, Meindl A, Schmutzler RK, et al. RAD51C deletion screening identifies a recurrent gross deletion in breast cancer and ovarian cancer families. *Breast Cancer Res* 2013;**15**: R120.
342. Gorski B, Debniak T, Masojc B, Mierzejewski M, Medrek K, Cybulski C, Jakubowska A, Kurzawski G, Chosia M, Scott R, Lubinski J. Germline 657del5 mutation in the NBS1 gene in breast cancer patients. *Int J Cancer* 2003;**106**: 379-81.
343. Cybulski C, Gorski B, Debniak T, Gliniewicz B, Mierzejewski M, Masojc B, Jakubowska A, Matyjasik J, Zlowocka E, Sikorski A, Narod SA, Lubinski J. NBS1 is a prostate cancer susceptibility gene. *Cancer Res* 2004;**64**: 1215-9.
344. Huzarski T, Cybulski C, Jakubowska A, Byrski T, Gronwald J, Domagala P, Szwiec M, Godlewski D, Kilar E, Marczyk E, Siolek M, Gorski B, et al. Clinical characteristics of breast cancer in patients with an NBS1 mutation. *Breast Cancer Res Treat* 2013;**141**: 471-6.
345. Xie ZM, Li LS, Laquet C, Penault-Llorca F, Uhrhammer N, Xie XM, Bignon YJ. Germline mutations of the E-cadherin gene in families with inherited invasive lobular breast carcinoma but no diffuse gastric cancer. *Cancer* 2011;**117**: 3112-7.
346. Ping Z, Siegal GP, Harada S, Eltoum IE, Youssef M, Shen T, He J, Huang Y, Chen D, Li Y, Bland KI, Chang HR, et al. ERBB2 mutation is associated with a worse prognosis in patients with CDH1 altered invasive lobular cancer of the breast. *Oncotarget* 2016;**7**: 80655-63.
347. Christgen M, Steinemann D, Kuhnle E, Langer F, Gluz O, Harbeck N, Kreipe H. Lobular breast cancer: Clinical, molecular and morphological characteristics. *Pathol Res Pract* 2016;**212**: 583-97.
348. Arpino G, Bardou VJ, Clark GM, Elledge RM. Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome. *Breast Cancer Res* 2004;**6**: R149-56.
349. Desmedt C, Zoppoli G, Gundem G, Pruneri G, Larsimont D, Fornili M, Fumagalli D, Brown D, Rothe F, Vincent D, Kheddoumi N, Rouas G, et al. Genomic Characterization of Primary Invasive Lobular Breast Cancer. *J Clin Oncol* 2016;**34**: 1872-81.
350. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, Bowlby R, Shen H, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 2015;**163**: 506-19.

351. Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM, Scott C, Weitzel JN, Oaknin A, Loman N, Lu K, Schmutzler RK, 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;**376**: 245-51.
352. Rottenberg S, Jaspers JE, Kersbergen A, van der Burg E, Nygren AO, Zander SA, Derksen PW, de Bruin M, Zevenhoven J, Lau A, Boulter R, Cranston A, et al. High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. *Proceedings of the National Academy of Sciences of the United States of America* 2008;**105**: 17079-84.
353. Mavaddat N, Rebbeck TR, Lakhani SR, Easton DF, Antoniou AC. Incorporating tumour pathology information into breast cancer risk prediction algorithms. *Breast Cancer Res* 2010;**12**: R28.
354. Sinilnikova OM, Dondon MG, Eon-Marchais S, Damiola F, Barjhoux L, Marcou M, Verny-Pierre C, Sornin V, Toulemonde L, Beauvallet J, Le Gal D, Mebirouk N, et al. GENESIS: a French national resource to study the missing heritability of breast cancer. *BMC Cancer* 2016;**16**: 13.
355. Osborne RH, Hopper JL, Kirk JA, Chenevix-Trench G, Thorne HJ, Sambrook JF. kConFab: a research resource of Australasian breast cancer families. Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. *Med J Aust* 2000;**172**: 463-4.
356. Balleine RL, Murali R, Bilous AM, Farshid G, Waring P, Provan P, Byth K, Thorne H, kConFab I, Kirk JA. Histopathological features of breast cancer in carriers of ATM gene variants. *Histopathology* 2006;**49**: 523-32.
357. Chenevix-Trench G, Spurdle AB, Gatei M, Kelly H, Marsh A, Chen X, Donn K, Cummings M, Nyholt D, Jenkins MA, Scott C, Pupo GM, et al. Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst* 2002;**94**: 205-15.
358. Waddell N, Jonnalagadda J, Marsh A, Grist S, Jenkins M, Hobson K, Taylor M, Lindeman GJ, Tavtigian SV, Suthers G, Goldgar D, Oefner PJ, et al. Characterization of the breast cancer associated ATM 7271T>G (V2424G) mutation by gene expression profiling. *Genes Chromosomes Cancer* 2006;**45**: 1169-81.
359. Campos CZ, Losi Guembarovski R, de Oliveira CEC, Banin Hirata BK, Vitiello GAF, Dias FL, Hiroki CH, Watanabe MAE, Mazzuco TL. Glutathione S-transferases deletions may act as prognosis and therapeutic markers in breast cancer. *Clin Exp Med* 2017.
360. Andrieu N, Easton DF, Chang-Claude J, Rookus MA, Brohet R, Cardis E, Antoniou AC, Wagner T, Simard J, Evans G, Peock S, Fricker JP, et al. Effect of chest X-rays on the risk of breast cancer among BRCA1/2 mutation carriers in the international BRCA1/2 carrier cohort study: a report from the EMBRACE, GENEPSO, GEO-HEBON, and IBCCS Collaborators' Group. *J Clin Oncol* 2006;**24**: 3361-6.
361. Andrieu N, Goldgar DE, Easton DF, Rookus M, Brohet R, Antoniou AC, Peock S, Evans G, Eccles D, Douglas F, Nogues C, Gauthier-Villars M, et al. Pregnancies, breast-feeding, and breast cancer risk in the International BRCA1/2 Carrier Cohort Study (IBCCS). *J Natl Cancer Inst* 2006;**98**: 535-44.
362. Li J, Meeks H, Feng BJ, Healey S, Thorne H, Makunin I, Ellis J, kConFab I, Campbell I, Southey M, Mitchell G, Clouston D, et al. Targeted massively parallel sequencing of a panel of putative breast cancer susceptibility genes in a large cohort of multiple-case breast and ovarian cancer families. *J Med Genet* 2016;**53**: 34-42.

Annexes

RESEARCH ARTICLE

ABRAXAS (FAM175A) and Breast Cancer Susceptibility: No Evidence of Association in the Breast Cancer Family Registry

Anne-Laure Renault¹, Fabienne Lesueur², Yan Coulombe³, Stéphane Gobeil¹, Penny Soucy¹, Yosr Hamdi¹, Sylvie Desjardins¹, Florence Le Calvez-Kelm⁴, Maxime Vallée^{1,4}, Catherine Voegelé⁴, The Breast Cancer Family Registry^{5,6,7,8,9,10†}, John L. Hopper⁵, Irene L. Andrulis^{6,7}, Melissa C. Southey⁸, Esther M. John^{9,10}, Jean-Yves Masson^{3,11}, Sean V. Tavtigian^{12,13}, Jacques Simard^{1*}

1 Genomics Center, Centre Hospitalier Universitaire de Québec Research Center and Laval University, Quebec, Canada, **2** INSERM, U900, Mines ParisTech, Institut Curie, Paris, France, **3** Genome Stability Laboratory, Centre Hospitalier Universitaire de Québec Research Center, HDQ Pavillon, Oncology Axis, Quebec, Canada, **4** Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France, **5** Center for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Victoria, Australia, **6** Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada, **7** Department of Molecular Genetics, University of Toronto, Toronto, Canada, **8** Genetic Epidemiology Laboratory, The University of Melbourne, Victoria, Australia, **9** Cancer Prevention Institute of California, Fremont, United States of America, **10** Stanford University School of Medicine and Stanford Cancer Institute, Stanford, United States of America, **11** Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University, Quebec, Canada, **12** Department of Oncological Sciences, University of Utah, Salt Lake City, United States of America, **13** Huntsman Cancer Institute, University of Utah, Salt Lake City, United States of America

† Complete membership of the BCFR can be found in the Acknowledgments.

* Jacques.Simard@crchudequebec.ulaval.ca



CrossMark
click for updates

OPEN ACCESS

Citation: Renault A-L, Lesueur F, Coulombe Y, Gobeil S, Soucy P, Hamdi Y, et al. (2016) *ABRAXAS (FAM175A) and Breast Cancer Susceptibility: No Evidence of Association in the Breast Cancer Family Registry*. PLoS ONE 11(6): e0156820. doi:10.1371/journal.pone.0156820

Editor: Alvaro Galli, CNR, ITALY

Received: November 12, 2014

Accepted: May 19, 2016

Published: June 7, 2016

Copyright: © 2016 Renault et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data were generated by the authors through mutation screening of genomic DNA samples obtained from the BCFR. Phenotypic data were also obtained from the BCFR along with the samples. The summary data is included in the manuscript however we do not have permission to make the individual raw data publicly available. There are ethics restrictions at the BCFR sites, whereby consent indicates that data can be shared with external investigators, but under a collaboration agreement. We are therefore obligated to submit these individual data to the BCFR Informatics Center where other researchers may then

Abstract

Approximately half of the familial aggregation of breast cancer remains unexplained. This proportion is less for early-onset disease where familial aggregation is greater, suggesting that other susceptibility genes remain to be discovered. The majority of known breast cancer susceptibility genes are involved in the DNA double-strand break repair pathway. *ABRAXAS* is involved in this pathway and mutations in this gene impair BRCA1 recruitment to DNA damage foci and increase cell sensitivity to ionizing radiation. Moreover, a recurrent germline mutation was reported in Finnish high-risk breast cancer families. To determine if *ABRAXAS* could be a breast cancer susceptibility gene in other populations, we conducted a population-based case-control mutation screening study of the coding exons and exon/intron boundaries of *ABRAXAS* in the Breast Cancer Family Registry. In addition to the common variant p.Asp373Asn, sixteen distinct rare variants were identified. Although no significant difference in allele frequencies between cases and controls was observed for the identified variants, two variants, p.Gly39Val and p.Thr141Ile, were shown to diminish phosphorylation of gamma-H2AX in MCF7 human breast adenocarcinoma cells, an important biomarker of DNA double-strand breaks. Overall, likely damaging or neutral variants were evenly represented among cases and controls suggesting

apply to obtain the data by initiating a request for collaboration through submission of a Concept form to the BCFR Administrative Coordinating Center (contact: Ms. Ly Ngo ly.ngo@cpic.org). BCFR is willing to share data and has done so extensively in the past. Guidelines for collaboration with the BCFR and application forms for data access are available at <http://www.bcfamilyregistry.org>.

Funding: This work was supported by the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program – grant # CRN-87521, the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701 and by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (GPH-129344), the Ministère de l'Économie, de la Science et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation for the PERSPECTIVE project, the National Cancer Institute (NCI) and the National Institute of Health (NIH) grant R01 CA121245. This work was also supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. J.S. is Chairholder of the Canada Research Chair in Oncogenetics. J-Y.M. is a FQRS chercheur national investigator and was supported by a Canadian Institutes of Health Research operating grant. I.L.A. is the Anne and Max Tanenbaum Chair in Molecular Medicine at the University of Toronto. 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.

that rare variants in *ABRAXAS* may explain only a small proportion of hereditary breast cancer.

Introduction

DNA damage induced by endogenous and exogenous genotoxic agents can promote genomic instability and directly lead to various diseases, particularly cancer. Functionally intact DNA double-strand break (DSB) repair machinery is essential for the maintenance of genomic integrity and stability. Mutations in genes coding for proteins involved in this pathway have been shown to cause chromosomal aberrations and defective cell cycle checkpoints, leading to tumor development [1]. Germline mutations in genes involved in homologous recombination DNA repair have been associated with breast cancer susceptibility. These include high-penetrance susceptibility genes such as *BRCA1* [2,3], *BRCA2* [4,5] and *TP53* [6] and the moderate penetrance genes *CHEK2* [7,8], *ATM* [9,10], *BRIP1* [11], *PALB2* [12,13] and *MRE11A*, *RAD50* and *NBN* [14]. However, despite technical progress and studies with greater statistical power, only approximately 35% of the familial relative risk of breast cancer is currently explained by the known high- and intermediate-risk genes, suggesting that other breast cancer susceptibility genes, possibly involved in the homologous recombination repair (HRR) pathway, remain to be discovered.

ABRAXAS (FAM175A) (MIM 611143; [NM_139076.2](#)) is a coiled-coil domain-containing protein that forms, along with Rap80, BRCC36, BRE and BABAM1, the A-Complex [15–17], which regulates the G2/M checkpoint and DNA-end resection during HRR [18–20]. Studies have reported that *ABRAXAS*-depleted cells have augmented levels of single-strand DNA and show increased binding of RPA and RAD51 proteins, indicative of increased resected DNA ends [20,21]. Excessive DNA resection can lead to loss of genomic integrity. *ABRAXAS* also directly interacts with the *BRCA1* BRCT (*BRCA1* carboxyl-terminal) repeats through its phosphorylated SPxF motif [22–24] and contributes to *BRCA1*-dependent DNA damage responses by localizing *BRCA1* to DNA damage foci [16]. Depletion of *ABRAXAS* impedes the recruitment of *BRCA1* to DNA damage sites, leading to impairment of G2/M checkpoint control after ionizing radiation (IR) induction [25].

Further attesting to the critical role of *ABRAXAS* in cellular response to DNA damage and potential involvement in cancer susceptibility, Castillo et al. recently reported that both homozygous and heterozygous *Abraxas* knockout mice exhibited decreased survival and increased tumor incidence [26]. Analysis of gene expression levels in human tumors in the TCGA database revealed reduced *ABRAXAS* gene expression in numerous cancers, including breast cancer. This study also showed that gene copy number loss of the *ABRAXAS* locus at chromosome 4q21 is frequently found in ovarian and breast cancers and that this loss is well correlated with reduced *ABRAXAS* expression levels in these cancers [26].

Solyom et al. reported the screening of 125 Northern Finnish breast cancer families for coding region and splice-site *ABRAXAS* mutations [27]. Their study identified a novel germline mutation (p.Arg361Gln) in *ABRAXAS* in three of the families. Segregation analysis was performed in two of the mutation-positive families, showing co-segregation between the p.Arg361Gln mutation and breast cancer phenotype. The missense mutation is located in the nuclear localization sequence (NLS) of *ABRAXAS* and affects the nuclear localization of the protein. Consequently, this mutation reduces the formation of *BRCA1* and Rap80 foci at DNA damage sites, leading to IR hypersensitivity of cells and partially impairing the G2/M

checkpoint [27]. Another recent study, reporting the sequencing of several homologous recombination genes in 390 ovarian carcinomas, identified the loss-of-function *ABRAXAS* germline mutation, c.1106insG, in 2 subjects, representing 2% of identified deleterious mutations [28].

Based on the evidence that a germline *ABRAXAS* mutation is associated with an increased risk of breast cancer in the Finnish population and owing to the crucial role of this protein in DSB repair, we sought to estimate the frequencies and nature of rare *ABRAXAS* variants in a sample of women with early onset breast cancer ($N = 1,332$) and frequency matched controls ($N = 1,123$) from three population-based centers of the Breast Cancer Family Registry (BCFR) [29] by screening *ABRAXAS* exons and exon/intron boundaries. We then followed a similar *in silico*-driven analysis strategy to that previously proven successful to identify the intermediate penetrance susceptibility alleles in *ATM* [10], *CHEK2* [7], *XRCC2* [30], *MRE11*, *RAD50* and *NBN* [14].

Materials and Methods

Ethics statement

This study was approved by the Research Ethics Committee of the Centre Hospitalier Universitaire de Québec (Project 123.05.08 / MP-CHUQ-CHUL-08-009), the University of Utah Institutional Review Board (IRB), the Ethics committee of the International Agency for Research on Cancer (IARC) and the local IRBs of the BCFR centers: the Health Sciences Human Ethics Subcommittee of the University of Melbourne, Australia; the Institutional Review Board of the Cancer Prevention Institute of California; and the Research Ethics Board of Mount Sinai Hospital, Ontario, Canada, from which samples were received. All participants gave written informed consent.

DNA samples

Subjects ($N = 2,455$) were selected from the population-based BCFR centers [29]. Women were recruited between 1995 and 2005 at three centers: the Cancer Prevention Institute of California (USA), the Cancer Care Ontario (Canada) and the University of Melbourne (Australia). Selection criteria for cases ($N = 1,332$) were: the age at breast cancer diagnosis (≤ 45 years), race/ethnicity, and grandparents' country of origin consistent with the ethnic heritage, that is Caucasian, Latino, East Asian or African-American ancestry. Cases were matched with controls ($N = 1,123$) within each center according to the age at recruitment (± 10 years from the age at diagnosis) and race/ethnicity. However, because of limited availability of controls in some racial/ethnic and/or age groups, the matching was not always one-to-one in these groups (Table 1).

Mutation screening

Mutation screening of the *ABRAXAS* gene was performed on Whole-Genome Amplified (WGA) DNA obtained by mixing 5 ng of amplified DNA from each of two independent WGA reactions. The 9 coding exons, exon-intron boundaries and part of the promoter region were screened by High Resolution Melting curve (HRM) analysis on a LightCycler 480 (Roche Applied Science, Indianapolis, USA), followed by direct Sanger sequencing of the samples for which an aberrant melting curve was indicative of the presence of a sequence variant (S1 Fig). All exons and part of the promoter region were amplified using specific primer pairs (S1 Table) in PCR reactions of 10 μ l: each reaction contained 10 ng of WGA DNA, 5 μ l of Master mix (Roche Applied Science, Indianapolis, USA), 1 μ l of primers (0.5 μ M each) and 1.2 μ l of $MgCl_2$ (3.6 μ M). The cycling conditions were as follow: 95°C for 5 min, 40 cycles of DNA denaturation

Table 1. Distribution of cases and controls by study center and by ethnicity in the BCFR.

Study center	Ethnic group	Cases (N = 1,332)	Controls (N = 1,123)	Total (N = 2455)
BCFR Ontario	All	314	463	777
	Caucasian	302	459	761
	East Asian	8	4	12
	Latino	4	0	4
	African-American Ancestry	0	0	0
BCFR Northern California	All	421	136	557
	Caucasian	0	0	0
	East Asian	177	54	231
	Latino	146	46	192
	African-American Ancestry	98	36	134
BCFR Australia	All	597	524	1121
	Caucasian	561	510	1071
	East Asian	28	13	41
	Latino	8	1	9
	African-American Ancestry	0	0	0

doi:10.1371/journal.pone.0156820.t001

at 95°C for 10 sec, annealing for 45sec at temperatures depending on each primer pair (S1 Table), and elongation at 72°C for 14 sec. According to the manufacturer’s recommendations, the size of amplicons was 250 bp or less in order to optimize amplification specificity and variant detection sensitivity. Prior to HRM analysis, a melting step consisting of a denaturation step at 95°C for 1 min, and a cooling step at 40°C for 1 min was carried out for heteroduplex formation. Finally, the PCR products were heated from 65°C to 95°C to determine the melting temperature of the amplicons. For each amplicon, samples showing an aberrant melting curve were purified on silica filtration microplates (Whatman hydrophilic GF/C Filter, GE Healthcare Life Science, Piscataway, USA) prior to Sanger sequencing in order to identify the variation causative of the observed difference in melting curves. Sequencing was performed using Big Dye Terminator chemistry on an ABI Prism 3730xl automated sequencer from Applied Biosystems (Applied Biosystems, Foster City, CA, USA).

All sequence variants that were either unreported or had an allele frequency of <1% in the large scale reference groups “Caucasian Americans”, “African Americans” and “East Asians” based on Exome Variant Server (EVS) [31] and 1,000 genomes project (1000G) data (1000 Genomes) were confirmed by concordant variant sequencing calls of the two independent re-amplified WGA reaction products. All samples that failed either at the PCR or sequencing reaction stage were re-amplified from WGA DNAs or genomic DNAs. Samples that still did not provide satisfactory mutation screening results for at least 80% of the ABRAXAS coding sequence were excluded from further analysis. A total of 22 subjects were excluded from the analyses (S2 Table).

Alignments and scoring of missense substitution

The multiple sequence alignment software M-Coffee, which is a part of the T-coffee software (<http://tcoffee.crg.cat>) (Tree-based Consistency Objective Function For alignment Evaluation) [32, 33] was used to prepare a protein multiple sequence alignment with ABRAXAS orthologs. Ten protein sequences of ABRAXAS were used, ranging from Human (*Homo sapiens*, NP_620775.2) to the most divergent zebrafish sequence (*Danio rerio*, NP_001005993.1), and including chimpanzee (*Pan troglodytes*, JAA36615.1), orangutan (*Pongo abelii*,

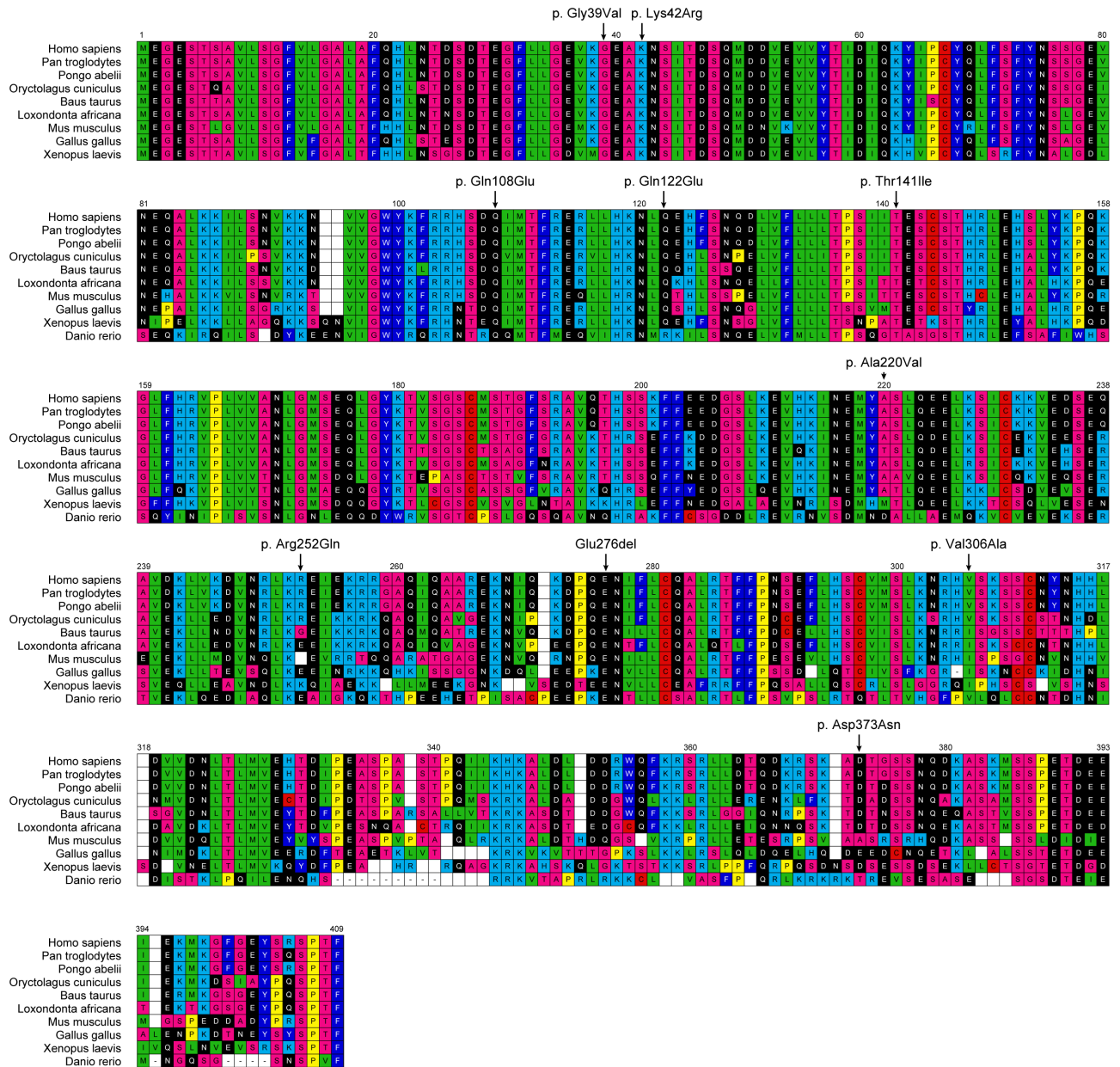


Fig 1. ABRAXAS multiple-sequence alignment. Substitution designations are indicated above the corresponding human reference sequence residue. Amino acid symbols are colored to represent standard Dayhoff groupings.

doi:10.1371/journal.pone.0156820.g001

ENSPPY00000016637), rabbit (*Oryctolagus cuniculus*, ENSOCUP00000011470), cow (*Bos taurus*, NP_0010015516.1), elephant (*Loxondonta africana*, ENSLAFP00000000891), mouse (*Mus musculus*, NP_765993), chicken (*Gallus gallus*, NP_001026315.2) and frog (*Xenopus laevis*, NP_001005339.1) (Fig 1 and S3 Table).

The impact of the missense substitutions identified during our ABRAXAS mutation screening was scored using our M-coffee alignment with the two prediction software programs Align-GVGD (<http://agvgd.iarc.fr>) and Sorting Intolerant From Tolerant (SIFT) (<http://sift.jcvi.org/>). PolyPhen2.1 (<http://genetics.bwh.harvard.edu/pph2/>) was also used with its

precompiled alignment. Briefly, Align-GVGD classifies missense variants in a query sequence into 7 grades, from the most deleterious C65 to the least deleterious C0, with the intermediate grades C15, C25, C35, C45 and C55 [34]. The program is based on Grantham calculation [35], a combination of Grantham Variation (GV) which measures the amount of observed biochemical evolutionary variation at a specific position of the alignment, and Grantham Deviation (GD) which measures the biochemical difference between the missense residue and the range of variation observed at this position in the alignment. SIFT is a sequence homology-based tool that predicts variants in the query sequence as “tolerated” or “deleterious”, calculating normalized probabilities for all possible substitutions in the multiple sequence alignment. Variants with normalized probabilities <0.05 are predicted to be “deleterious” and those ≥ 0.05 are predicted to be “tolerated” [36]. PolyPhen2 classifies variants as “benign”, “Possibly Damaging” and “Probably Damaging”, according to eight sequenced-based and three structure-based predictive features. We used the PolyPhen2 precompiled alignment constructed and refined through a pipeline that selects homologous sequences, among which orthologs and paralogs that may or may not be full length [37].

In silico analysis of variants

Intronic and exonic variants located in the vicinity of splice junction consensus sites were scored for their potential impact on splicing using the Splicing Prediction Module of Alamut (Interactive Biosoftware) which computes the following prediction methods: SpliceSiteFinder-like [38], MaxEntScan [39], NNSPLICE [40], GeneSplicer [41] and Human Splicing Finder [42]. Variants located near the beginning of the first exon were analyzed using the MatInspector software of Genomatix [43] to assess whether they altered transcription factor binding sites.

Statistical analysis

To assess the relationship between ABRAXAS variants and breast cancer risk we compared frequency distribution of rare missense substitutions and in-frame deletions in cases and in controls. The case-control statistical analysis has been described in detail previously [7]. Briefly, a single table with one entry per subject, zero or one rare sequence variant per subject, annotations for type of sequence variants, study center, case-control status, and race/ethnicity was constructed. Analyses were performed using the chi square test, the Fischer test and multivariable unconditional logistic regressions using STATA version 11 software (StataCorp, College Station, TX, USA). Differences in the case/control ratio between racial/ethnic groups and study center were accounted for by including categorical variables for each racial/ethnic group and each study center. As demonstrated previously, inclusion of interactions between study center and ethnic group did not change the estimates [7]. Adjustment for racial/ethnic group should capture confounding of genetic and social factors with interaction terms, allowing that this confounding effect may be different for the broadly labeled racial/ethnic groups in different centers.

Site-directed mutagenesis

To enable immuno-fluorescence assays for p.Gly39Val and p.Thr141Ile variants, site-directed mutagenesis was performed on a pOZ-ABRAXAS_cDNA-HA-Flag vector (kindly provided by Dr Roger A. Greenberg, University of Pennsylvania, Philadelphia, PA, USA), containing the full-length cDNA fragment encoding the complete amino acid sequence for human ABRAXAS. The vector was mutated at two positions, c.116G>T (p.Gly39Val) and c.422C>T (p.Thr141Ile) using the QuickChange Site-directed mutagenesis kit from Stratagene (Stratagene Cloning Systems, La Jolla, CA, USA) according to the manufacturer’s protocol and in

conditions specific to each primer pairs for variants c.116G>T (F: 5' -TTGGGGAAGTAAAA GTTGAAGCCAAGAACAG-3' /R: 5' -CTGTTCTGGCTTCAACTTTTACTTCCCAA-3') and c.422C>T (F: 5' -ACCAAGTATAATAATAGAAAAGCTGCTCTACTC-3' /R: 5' -GTAGAGCA GCTTTCTATTATTATACTTGGT-3') respectively. DNA templates were degraded with 1 μ L of DpnI at 37°C during 1 h. The resulting constructs were purified (Sigma-Aldrich Kit, St. Louis, USA) and insertion of the desired mutation was confirmed by direct sequencing of both strands using Big Dye Terminator chemistry on an ABI Prism 3730xl automated sequencer from Applied Biosystems (Applied Biosystems, Foster City, CA, USA).

Cell culture

The Human Embryonic Kidney cell line (HEK293T) was grown in DMEM high glucose (Wisent, St-Bruno, Canada) containing 10% FBS, 2 mM L-glutamine and penicillin/streptomycin. The ER+ human breast adenocarcinoma cells, MCF7, were cultured in DMEM/F12 (Wisent Bioproducts, St-Bruno, Quebec, Canada) supplemented with 5% FBS, 2 mM L-glutamine, penicillin/streptomycin, 15 mM HEPES, 0,2% sodium bicarbonate and 10E-9 M estradiol. The human cervical carcinoma cells, HeLa, were grown in DMEM medium (Wisent) with 5% FBS, 2 mM L-glutamine and penicillin/streptomycin. Cells were grown at 37°C in a 5% CO₂ incubator.

RNA interference

Two *ABRAXAS* short hairpin RNAs (shRNA) (designated shABX139 and shABX145) cloned into the pLKO.1-puro vector were retrieved from the Sigma Mission human shRNA library available at the CHU de Quebec Research Centre (clone numbers TRCN0000139032 and TRCN0000145012). The pLKO.1-puro vector encoding a scramble sequence not matching any mammalian sequence was used as a control (designated shNSCTL). Viral supernatants were generated by transfecting HEK293T cells with the shRNA constructs and the packaging vectors psPAX2 and pMD2.G (Addgene, Cambridge, MA). The high-titer lentiviral supernatants in the presence of 10 μ g/ml of polybrene were used to transduce MCF7 and HeLa cell lines. Two days later, puromycin (2 μ g/ml) was added to the culture media to select stably transduced cells. The expression levels of *ABRAXAS* were thereafter confirmed by quantitative PCR (see [S2 Fig](#) for knockdown efficiencies in MCF7 and HeLa cells).

Quantitative RT-PCR

Total RNA was extracted using TriReagent (Sigma-Aldrich) following manufacturer's instructions. Reverse transcription was performed with the M-MuLV Reverse Transcriptase (NEB) and a mix of random hexamers and anchored oligo(dT). Quantitative PCR reactions were performed using four ng of cDNA, analyzed in three replicates for each data point, in a quantitative PCR (CFX384 RealTimeSystem (Bio-Rad)) using the iTAQ Universal SYBR Green Supermix (Bio-Rad). The following forward and reverse primers were used: 5' -AGAGATAATTTAAAGTTC TTGACAAAAC-3' and 5' -TCAAATAATGGGTAAGAAAGAATAC-3' . PCR volume was 10 μ l and conditions were as follow: Initial activation at 95°C, 30s followed by 40 cycles at 95°C, 5s and 60°C, 30s; final cycle (melting curve) 65°C, 0.05s and 95°C 0.5s. Relative expressions were calculated using the CFX Manager Software v3.1 using U6 and ACTB as normalizers.

Immunofluorescence assays

Four hundred nanograms of empty pOZ-HA-Flag vector, pOZ-*ABRAXAS*-HA-Flag vector mutated at positions c.116G>T (p.Gly39Val), c.422C>T (p.Thr141Ile) or WT, were transfected in the two independent *ABRAXAS* KD MCF7 and HeLa cell lines (shABX139 and shABX145)

as well as in the shNSCTL MCF7 and HeLa cell lines with Effectene transfection agent (QIAGEN, Valencia, USA). Transfected MCF7 cells were treated with 50 ng/ml of neocarzinostatin (SIGMA) and transfected HeLa cells with 25ng/ml for 30 min and released in new media for 4 hours (20 hours for empty pOZ-HA-Flag vector specifically), washed twice in PBS, fixed with 2% paraformaldehyde in PBS for 10 min, washed with TBS and fixed with cold methanol (-20°C) for 5 min. Next, cells were washed once in TBS, permeabilized 5 min with PBS (0.2% Triton X-100) and washed three times 5 min with TBS. Then, cells were quenched with 0.1% Sodium Borohydride 5 min, washed once with TBS and blocked in PBS (10% goat serum and 1% BSA) 1h. Cells were then incubated 1 h with the primary antibody diluted in 1% BSA/TBS. Cells were washed three times 5min with TBS and incubated 1 h with the appropriate secondary antibody (1% BSA/TBS) conjugated to a fluorophore (Alexa Fluor 488 (green) and Alexa Fluor 568 (red) from Life Technologies). Cells were washed three times 10 min with TBS and coverslips were mounted onto slides with PBS-glycerol (90%) containing 1 mg/ml paraphenylenediamine and 0.2 mg/ml of 4,6-diamidino-2-phenylindole (DAPI). For gamma-H2AX scoring, images were obtained using a Leica CTR 6000 microscope. Number of foci per cell was automatically counted following background subtraction and deconvolution using Volocity software v 5.5 (Perkin-Elmer Improvision). Foci were scored according to intensity within a $\leq 0.7\mu\text{M}$ radius. P-values were obtained using Wilcoxon's Test with $N = 100$ cells from four independent experiments for mutant and WT pOZ-HA-Flag vectors, while $N = 129$ cells from three independent experiments were used to calculate p-values for empty pOZ-HA-Flag vector.

Transient transfection and transcriptional activity assay

Constructs: To assess the impact of variant p.21G>A (rs145796091) on transcriptional activity, a fragment containing exon 1 and part of the promoter of *ABRAXAS* was amplified using genomic DNA from an individual heterozygous at this position, and subcloned into the firefly luciferase-reporter pGL3-Basic vector (Promega, Madison, WI, USA). The following primers were used for amplification (5' - **CTAGCTAGCTAG**GTGGCATATCCACTGTGGCATCGT-3', 5' - **CCGCTCGAGCGG**AGGGCTAATGCTGGAGAAGACTTCGTGG-3'). The resulting constructs were sequenced to confirm the presence of the expected polymorphic site, amplified and then purified using Sigma GenElute HP Plasmid Kit (Sigma-Aldrich, St-Louis, MO, USA) prior to transfection. Three different clones were obtained for each genotype, in order to take into account possible inter-clone variability during luciferase reporter assays.

MCF7 cells were seeded in 24-well culture dishes at a density of 130,000 cells/well for 24 h prior to transfection. Transient transfection was performed using Lipofectamine 2000 transfection reagent (Life Technologies Inc., Ontario, Canada) according to the supplier's protocol. Briefly, MCF7 cells were co-transfected with 1 μg of pGL3-promoter genotype-specific constructs encoding a modified firefly luciferase gene and 10 ng of a CMV-driven Renilla luciferase pRL-CMV plasmid (Promega, Madison, USA) (ratio 100:1) to control for transfection efficiency. The promoterless pGL3-basic vector and pGL3-SV40 control vector, containing the SV40 early promoter, were used as negative and positive controls, respectively. Cells were harvested 24 h post-transfection and luciferase reporter gene activities measured with the Dual-Luciferase Reporter Assay System according to the manufacturer's instructions (Promega, Madison, USA) in a M-1000 luminometer (TECAN). The promoterless pGL3-basic vector was used to measure basal expression levels. Each experiment was performed three times. Luciferase activity was normalized to Renilla luciferase. Data from four replicates per construct were analyzed with a mixed model including the fixed effect of genotype and the random effects of experiment number and clone within experiment number. The analysis was done at the 0.05 level of significance. The model was fitted using the lme function of the nlme R package [44].

Splicing reporter mini-gene assays: Vector construction, RNA extraction and RT-PCR

Exon 1, including the sequence variant c.21G>A (p.Ser7Ser), exon 2 and exon 3 of *ABRAXAS* were amplified separately by PCR using 50 ng of genomic DNA from an individual carrying the variant, and specific primer pairs (S4 Table). The amplified products were subcloned in a pcDNA3.1(+) vector (Invitrogen, Life Technologies, Burlington, Canada) in a three-step strategy: first, exon 1 was digested with NheI and HindIII (New England Biolabs, Beverly, MA) and inserted in the vector, then exon 2 was digested with HindIII and EcoRV and inserted in the pcDNA3.1-exon1, and finally exon 3 was inserted following its digestion with EcoRV and XhoI. The obtained pcDNA3.1-exon1-exon2-exon3 construction was transfected in HEK293T cells using Lipofectamine 2000 (Invitrogen, Life Technologies, Burlington, Canada) according to the supplier's protocol. After 24 hours, cells were collected and total RNA was extracted using the TRI-Reagent Solution Protocol according to supplier's instructions (Sigma-Aldrich, St Louis, USA). Reverse transcription polymerase chain reaction (RT-PCR) was performed using a PROMEGA Access RT-PCR System kit (PROMEGA, Madison, USA) according to the manufacturer's protocol, with 50ng of the total extracted RNA and the primer pair F: 5' - ACGACTCACTATAGGGACCACAGG-3' / R: 5' -AGAAGGCACAGTCGAGGCTGATCAG-3' specific to the exogenous mRNA of the construction. Finally, the RT-PCR product was analyzed both by gel-electrophoresis, using 2% agarose gel, and by Sanger sequencing.

Results

Case-control mutation screening

A total of 2,455 subjects from the three study centers that constitute the population-based arm of the BCFR were screened for mutations. The distribution of cases and controls by race/ethnicity and study center are detailed in Table 1. In addition to the common missense substitution p.Asp373Asn (rs13125836), HRM screening revealed sixteen rare distinct sequence variants, including an in-frame deletion found in one case and one control, eight missense substitutions, four silent substitutions, two intronic variations and one variant located in the 5'-UTR region (S1 Fig). The distribution of the rare variants (minor allele frequency less than 1% in Exome Variant Server) in cases and controls is shown in Table 2.

Analysis of missense substitutions and in-frame indels

The potential functional impact of the missense substitutions was assessed using the three *in silico* prediction programs: Align-GVGD, SIFT and PolyPhen2. The common SNP p.Asp373Asn and the rare missense substitutions p.Lys42Arg, p.Gln122Glu, p.Ala220Val and p.Arg252Gln were predicted to be benign by at least two prediction tools, while p.Gly39Val, p.Gln108Glu, p.Thr141Ile and p.Val306Ala were predicted to be damaging or possibly damaging (Table 2). In particular, the protein multiple sequence alignment revealed complete evolutionary conservation of the ancestral amino acid sequence in the regions surrounding codon 39 and codon 141 among all species investigated (Fig 1), and p.Gly39Val and p.Thr141Ile were assigned to the most severe grade with the three algorithms.

A simple binary classification combining all rare variants affecting the coding sequence of *ABRAXAS* did not reveal any significant difference between cases and controls ($p = 0.27$). A similar result was obtained when excluding the four likely neutral variants graded C0 with Align-GVGD (Table 3). Because the missense substitution p.Thr141Ile appeared to have a frequency >1% in the European population in our study, we assessed its contribution to breast cancer risk independently and ruled out an association with the disease in the BCFR population (Table 4).

Table 2. Distribution of ABRAXAS rare variants (i.e. with a minor allele frequency <1% in the Exome Variant Server (EVS)) identified in the BCFR.

Variant [¶]	Effect on protein	Reference	Cases (N = 1,318)	Controls (N = 1,115)	Prediction of variant effect		
					Align-GVGD	SIFT	PolyPhen2
Missense substitutions							
c.116G>T	p.Gly39Val	-	1	0	C65	Damaging	Probably Damaging
c.125A>G	p.Lys42Arg	rs201948472	1	1	C0	Tolerated	Probably Damaging
c.322C>G	p.Gln108Glu	-	1	1	C25	Damaging	Possibly Damaging
c.364C>G	p.Gln122Glu	rs137876115	3*	0	C0	Tolerated	Possibly Damaging
c.422C>T	p.Thr141Ile	rs150207999	17	13	C65	Damaging	Probably Damaging
c.659C>T	p.Ala220Val	-	1	1	C0	Tolerated	Possibly Damaging
c.755G>A	p.Arg252Gln	rs114513239	3	2	C0	Tolerated	Benign
c.917T>C	p.Val306Ala	rs138986552	1	0	C25	Damaging	Benign
In-frame deletion							
c.826-828delAGG	p.Glu276del	-	1	1	N/A	N/A	N/A
Silent substitutions							
c.21G>A	p.Ser7Ser	rs145796091	17*	4*	N/A	N/A	N/A
c.33G>C	p.Ser11Ser	-	5	0	N/A	N/A	N/A
c.951C>T	p.Leu317Leu	rs79357787	2	0	N/A	N/A	N/A
c.1128T>C	p.Ser376Ser	-	1	1	N/A	N/A	N/A
5'UTR substitution							
c.-4T>C		rs202166386	1	2*	N/A	N/A	N/A
Intronic variations							
c.179-35_179-32delTAAT	-	rs199678739	31*	26*	N/A	N/A	N/A
c.681+21C>T	-	rs188169329	0	1	N/A	N/A	N/A

N/A, Not Applicable.

[¶] NM_139076.2 was chosen as reference sequence.

*One Caucasian control carried both c.-4T>C and c.179-35_179-32delTAAT. One Latino control and one Latino case carried both c.21G>A and c.179-35_179-32delTAAT. One Latino case carried both c.364C>G and c.179-35_179-32delTAAT.

doi:10.1371/journal.pone.0156820.t002

Table 3. Analysis of potentially pathogenic ABRAXAS in-frame deletion or rare missense substitutions.

Class	Cases (N)	Controls (N)	Crude OR (95% CI)	Adj ^a OR (95% CI) (ethnicity and center)
Non-carriers	1,289	1,096		
All rare variants (incl. C0)	29	19	1.30 (0.72, 2.33) p = 0.38	1.41 (0.77, 2.57) p = 0.27
All rare variants (>C0) ^b	21	15	1.19 (0.61, 2.32) p = 0.61	1.32 (0.67, 2.63) p = 0.42
All rare variants (incl. C0), excluding p.Thr141Ile	12	6	1.70 (0.64, 4.55) p = 0.29	1.61 (0.58, 4.47) p = 0.36

^a OR are adjusted for race or ethnicity (Caucasian, East Asian, African American or Latina) and study center.

^b In the binary analysis, only carriers of a missense substitution with grade>C0 or of an in-frame deletion (IFR) were considered.

doi:10.1371/journal.pone.0156820.t003

Table 4. Distribution of p.Thr141Ile, p.Ser7Ser and p.Ser11Ser by race/ethnicity.

Variant	Race/Ethnicity	Cases	Controls	Trend test p-value (crude analysis)	Trend test p-value (adjusted analysis) ^a
p.Thr141Ile	All	17/1301	13/1102	0.78	0.49
	European	13/844	13/949	0.77	0.76
	Latino	2/151	0/47	-	-
	East Asian	1/209	0/71	-	-
	African-American ancestry	1/97	0/35	-	-
p.Ser7Ser	All	17/1301	4/1111	0.021	0.61
	European	0/857	0/962	-	-
	Latino	12/141	4/43	0.88	0.90
	East Asian	5/205	0/71	(0.33)*	-
	African-American ancestry	0/98	0/35	-	-
p.Ser11Ser	All	5/1313	0/1115	(0.07)*	-
	European	0/857	0/962	-	-
	Latino	0/153	0/47	-	-
	East Asian	5/205	0/71	(0.33)*	-
	African-American ancestry	0/98	0/35	-	-

*Fisher's exact test p-value

^a OR are adjusted for study center (and for race or ethnicity in the combined analysis).

doi:10.1371/journal.pone.0156820.t004

We also analyzed independently the common missense substitution p.Asp373Asn (rs13125836) in this BCFR series. The observed minor allele (A) frequency in the four racial/ethnic groups represented is shown in Table 5. No departure from Hardy-Weinberg equilibrium was observed in either cases or controls. Overall, there was no significant difference in allele frequencies between cases and controls for this SNP when pooling the different populations or after stratifying by race/ethnicity.

Immunofluorescence assays of ABRAXAS missense substitutions graded damaging by Align-GVGD, SIFT and PolyPhen2 algorithms

To assess the impact of the two variants classified as damaging (C65; Align GVGD) for protein function (p.Thr141Ile and p.Gly39Val), formation of repair foci was analyzed by

Table 5. Stratified analyses of the common SNP rs13125836 (c. 1117G>A, p.Asp373Asn) on breast cancer risk in the BCFR.

	Number of genotyped subjects Cases / Controls	A allele frequency Cases / Controls	Chi ² P-value ^a	Log-additive model ^b OR* [95% CI]	P-trend
All	1,318 / 1,115	0.038 / 0.036	0.70	0.90 [0.65, 1.24]	0.50
By race/ethnicity					
European	857 / 962	0.035 / 0.040	0.48	0.85 [0.59, 1.21]	0.36
East Asian	210 / 71	0.017 / 0.0	0.12	-	-
Recent African ancestry	98 / 35	0.066 / 0.071	0.88	0.92 [0.30, 2.79]	0.88
Latina	153 / 47	0.049 / 0.043	0.80	0.85 [0.26, 2.83]	0.80

^aTest for the difference in A allele frequency between cases and controls.

^bResults of the logistic regression assuming a log-additive model with study center included in the regression model as covariate in the combined analysis, and with race/ethnicity and study center as covariates in the stratified analysis (*OR is given for heterozygous carriers of the A allele).

doi:10.1371/journal.pone.0156820.t005

immunofluorescence (IF) assays. To do this, we first generated MCF7 and HeLa *ABRAXAS* knockdown cells using two different shRNAs (see [S2 Fig](#) for expression knockdown efficiencies) to evaluate whether the knock down of *ABRAXAS* had a functional consequence. Using staining of gamma-H2AX as a marker for damaged DNA, we found that sh*ABRAXAS* cell lines (shABX139 and shABX145) showed a higher percentage of cells containing gamma-H2AX foci, compared to shNSCTL, 20 hrs post-treatment with the DSB-generating agent neocarzinostatin, indicative of a deficiency in DNA repair ([S3 Fig](#)). A significant difference is observed between both sh*ABRAXAS* clones and the shNSCTL in HeLa cells (p-values for shABX139 and shABX145 are 0.0005 and 0.0065, respectively). In the MCF7 cell line, a significant difference is observed for shABX145 (p-value = 0.00167) while an increase in the percentage of cells containing gamma-H2AX foci is also observed for shABX139, albeit this increase is not significant (p-value = 0.2145). Although these results appear to show that the partial knock down of *ABRAXAS* was sufficient to result in a functional consequence, these results should be interpreted with caution.

Thereafter, sh*ABRAXAS* cells were complemented with *ABRAXAS*-HA-Flag, *ABRAXAS*-S-HA-Flag p.Thr141Ile, or *ABRAXAS*-HA-Flag p.Gly39Val, to assess the impact of the two variants on DNA repair. As illustrated in [Fig 2A](#), IF of wild-type, p.Thr141Ile and p.Gly39Val *ABRAXAS* show that DNA repair foci are formed even in the presence of the variants, indicating that these variants do not negatively impact the recruitment of the A-complex to DNA damage sites. Next, we monitored the formation of gamma-H2AX foci after treatment with neocarzinostatin. After a 4h release, gamma-H2AX foci formation was significantly reduced in both sh*ABRAXAS* MCF7 cell lines (shABX139 and shABX145) transfected with p.Thr141Ile (p-value 2.01×10^{-9} for shABX139 and 1.62×10^{-9} for shABX145) and p.Gly39Val (p-value 4.39×10^{-14} for shABX139 and 5.26×10^{-10} for shABX145) compared to WT. On the other hand, we did not observe a significant difference between the non-silencing control cell line transfected with either the WT or the mutant constructs. Thus the significant difference in foci distribution observed in sh*ABRAXAS* MCF7 cell lines between variants p.Thr141Ile and p.Gly39Val and WT ([Fig 2B](#)) suggests that the two variants either: 1) impair formation of DNA repair foci or 2) impair DNA damage signaling after DSB formation. Experiments were also performed in the HeLa cell line ([S4 Fig](#)). Results did not show significant differences in the formation of gamma-H2AX foci between WT and p.Gly39Val in both sh*ABRAXAS* knock down cell lines (p-values 0.085 and 0.081 for clones shABX139 and shABX145 respectively), while a significant decrease was observed for p.Thr141Ile, albeit this difference being weaker than that observed in MCF7 cells (p-values 0.001 and 0.0005 for clones shABX139 and shABX145 respectively, compared to $p = 2 \times 10^{-9}$ and $p = 1.6 \times 10^{-9}$ in MCF7).

Analysis of silent substitutions, 5'-UTR and intronic variations

The variant c.-4T>C located in the 5'-UTR of *ABRAXAS* and the two silent substitutions c.21G>A (p.Ser7Ser) and c.33G>C (p.Ser11Ser) located near the beginning of the first exon were analyzed using the MatInspector software of Genomatix [43] in order to predict whether they altered transcription factor consensus binding sites. These analyses showed that the variant c.21G>A (p.Ser7Ser) abolished a consensus binding site for the Early Growth Response (EGR2) transcription factor, while the two other examined variants were not predicted to affect the binding sites of any transcription factors known to be involved in breast cancer. Gene reporter assays were performed in the breast cancer cell line MCF7 using a promoter fragment containing c.21G>A (p.Ser7Ser) to assess the impact of the variant on transcriptional activity. A 1.4 fold increase in transcriptional levels was observed for the construct containing the variant compared to the wild-type construct ($F_{(1,48)} = 45.6$, $p = 1.76 \times 10^{-8}$) ([Fig 3A](#)). Due to the

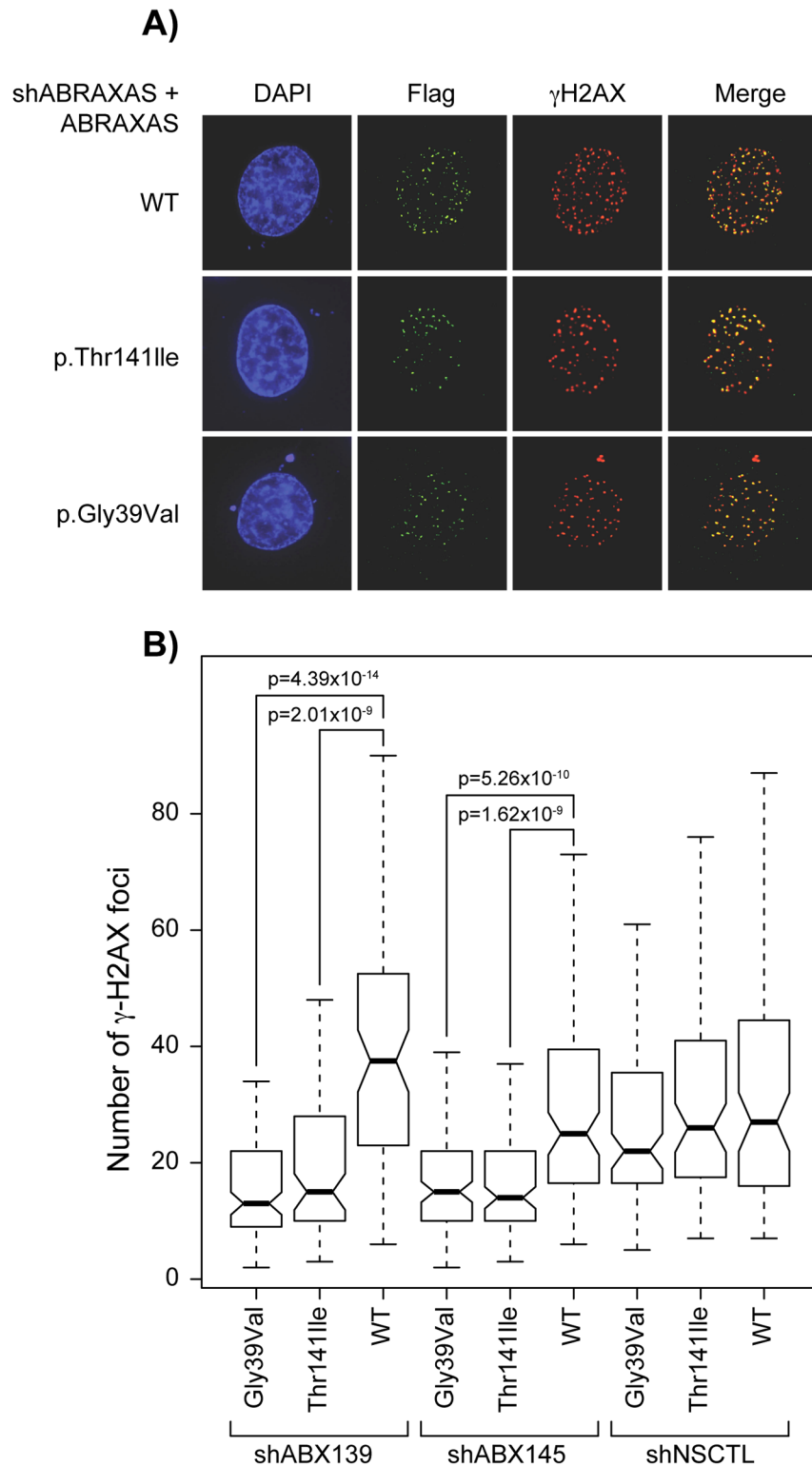


Fig 2. p.Gly39val and p.Thr141Ile ABRAXAS mutants have defects in gamma-H2AX formation. (A) Typical DNA damage foci of ABRAXAS in shABRAXAS (shABX145) MCF7 cells complemented with ABRAXAS-HA-Flag, ABRAXAS-HA-Flag pThr141Ile, or ABRAXAS-HA-Flag pGly39Val. The anti-Flag antibody was used to monitor ABRAXAS foci formation (green), anti-gamma-H2AX (red) and the merge picture is depicted. In blue, DAPI staining. (B) Quantification of gamma-H2AX foci formation in MCF7 cells

after neocarzinostatin treatment and release. P-values were obtained with a Wilcoxon's Test with N = 100 cells from four independent experiments.

doi:10.1371/journal.pone.0156820.g002

proximity of c.21G>A (p.Ser7Ser) variant to the beginning of the first exon, we also investigated whether this variant could alter the splicing of intron 1 using the *in silico* prediction program Human Splicing Finder [36]. Predictions using this program revealed that the variant disrupted the consensus binding site for hRNP8G9, a protein involved in the splicing machinery. In attempt to confirm this, splicing reporter mini-gene assays were used to evaluate the effect of this variant on splicing between exons 1 and 2 (Fig 3B). After transient transfection in HEK293T cells, the splicing patterns of the transcripts generated from the wild-type and c.21G>A variant constructs were compared by reverse transcription-PCR analysis and direct sequencing (Fig 3C). No differences in splicing patterns were observed suggesting that the variant has no impact on the splicing between exon 1, exon 2 and exon 3.

In silico analysis of the intronic variants c.681+21C>T and c.179-35_179-32delTAAT did not reveal any predicted impact on splice junction consensus sites.

Discussion

To our knowledge, this study is the largest case-control mutation screening investigating whether rare sequence variations within *ABRAXAS* contribute to breast cancer susceptibility. This gene was an appealing candidate because of (i) its central role in the formation of the A-complex and its critical functions during HR repair, (ii) its implication in *BRCA1* recruitment to DNA double-strand breaks, (iii) its interaction with the proteins encoded by the two known breast cancer susceptibility genes *BRCA1* and *BABAM1* (*MERIT40*) and (iv) its identification as a breast cancer susceptibility gene in Finnish breast cancer families.

Our mutation screening identified 17 *ABRAXAS* variants: 11 were already reported in the public databases such as EVS (Exome Variant Server) [31] and 1,000G (1000 Genomes) and 6 are novel variants. While the two missense variants, p.Thr141Ile (rs150207999) [45] and p.Asp373Asn (rs13125836) [27, 45] were previously reported in mutation screening studies performed in Spanish and Finnish high-risk breast cancer families, their respective frequencies in cases and controls in the previously mentioned studies [27, 45] did not suggest a role of these alleles in breast cancer susceptibility. Similarly, the frequencies of these variants in our study were not indicative of significant increased risk of breast cancer. However, *in silico* analyses using the three prediction tools Align-GVGD, SIFT and PolyPhen2 have classified the p.Thr141Ile variant at the highest grades of damage for the protein and this variant is located in the RAP80-binding domain. RAP80 recognizes the ubiquitinated proteins at sites of DNA breaks, which in turn facilitates the recruitment of *BRCA1* to DNA double strand breaks via a direct interaction between *BRCA1* and *ABRAXAS* [23, 46, 47]. Thus, an alteration of the sequence in this binding domain could impair the localization of RAP80 and other binding partners at sites where DNA repair is needed.

The second variant predicted to be deleterious to the protein, p.Gly39Val, was not previously reported and was only observed in one African-American case, therefore preventing the analysis of any association with breast cancer risk.

We thus investigated through immunofluorescence assays whether the two variants predicted to be most damaging, p.Thr141Ile and p.Gly39Val, had a functional impact on the protein (Fig 2). In MCF7 sh*ABRAXAS* cells complemented with the variants, both variants localized to DNA damage sites, however the number of gamma-H2AX foci formed was significantly different from WT complemented sh*ABRAXAS* cells. Taken together, these results suggest that these variants affect the DNA damage response and are thus likely to play a role in

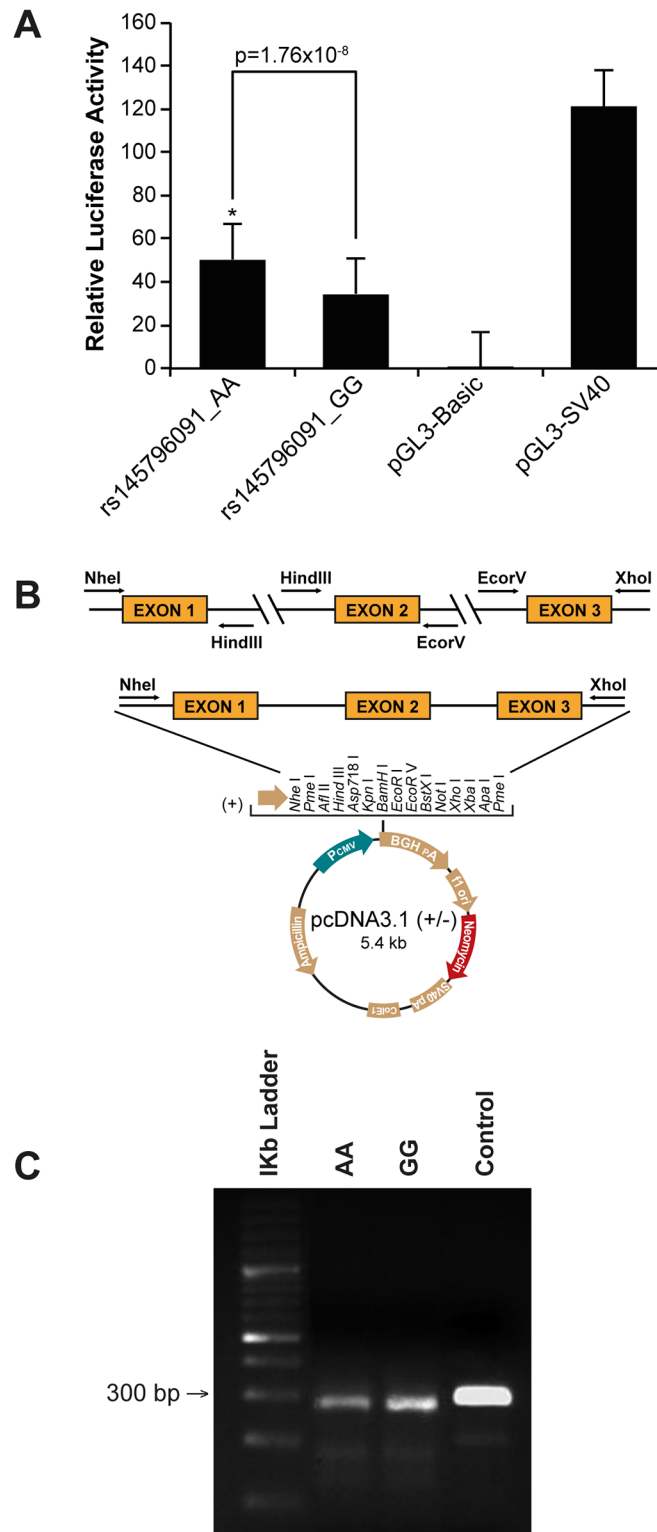


Fig 3. Functional assays assessing the impact of variant rs145796091 c.21G>A (p.Ser7Ser) on transcriptional activity and splicing efficiency in the MCF7 breast cancer cell line. A) Gene reporter assays. Luciferase activity was normalized to Renilla luciferase. Each experiment was performed three times. Data from four replicates per construct were analyzed with a mixed model including the fixed effect of genotype and the random effects of experiment number and clone within experiment number. B) Splicing

reporter mini-gene assays. Mini-gene constructions: exons 1, 2 and 3 of *ABRAXAS* were subcloned in a pcDNA3.1 vector. C) Gel-electrophoresis of mini-gene RT-PCR products on 2% agarose gels.

doi:10.1371/journal.pone.0156820.g003

breast cancer susceptibility, but much larger studies would be required to test this hypothesis due to the rarity of the two variants. The decrease of gamma-H2AX phosphorylation could be explained by several means including defective activation of ATM/ATR. Further work is warranted to confirm the functional impact of these variants on *ABRAXAS*.

Solyom *et al* provided evidence that a mutation in *ABRAXAS* could indeed have a significant impact of the proper functioning of the DNA repair pathway. It has recently been reported that the mutation c.1082G>A (p.Arg361Gln), identified in 2.4% of the 125 Finnish high-risk breast cancer families ($P = 0.002$ –familial cases versus controls), which impaired the localization of the protein in cultured cells, caused hypersensitivity to IR and reduced BRCA1 localization at sites of DNA damage [27]. This variant was not observed in our BCFR study population which was ascertained in North America and Australia, and included Caucasians of European Ancestry, Latinas and African-Americans suggesting that this variation may be specific to the Finnish population.

Analysis of silent and intronic substitutions did not reveal significant associations of these variants with breast cancer risk (Table 2). Interestingly, c.21G>A; p.Ser7Ser and c.33G>C; p.Ser11Ser were observed only in Latinas and East Asians (Table 4). The *in silico* analyses and experimental assays performed in this study did not support a functional role for these variants. Over-representation of these variants in these specific populations would warrant further investigation in these racial/ethnic groups using larger sample sizes, in order to provide statistical power to robustly detect an association with the disease.

Although the results of our study did not provide evidence that rare variants in *ABRAXAS* are associated with increased breast cancer risk in the populations studied, we cannot rule out the possibility that rare mutations in *ABRAXAS* may be involved in some high-risk families with more specific phenotypes. Solyom *et al*. reported that in addition to breast cancer, the families with p.Arg361Gln displayed some relatively rare types of cancer, such as lung and lip cancer and lymphoma of the throat [27]. Another study, reporting the sequencing of several homologous recombination genes in 390 ovarian carcinomas, identified the *ABRAXAS* germline mutation, c.1106insG, in 2 subjects, representing 2% of identified deleterious mutations [28]. Moreover, a recent genome-wide association study associated a common SNP (rs1494961) located downstream of the *ABRAXAS* gene with upper aero-digestive tract cancer risk (OR = 1.12, 95% CI = 1.08–1.17, $p = 1 \times 10^{-8}$) [48]. In view of this, much larger case-control studies are needed to determine whether allelic variants in this gene could be associated with a low or modest risk of breast cancer.

Supporting Information

S1 Fig. Mutation screening of *ABRAXAS* by High Resolution Melting curve analysis. Representative melting curves obtained from 384 samples, for exons where variants were observed. Panels (A) and (B) First exon, (C) Exon 2, (D) Exon 3, (E) and (F) Exon 5, (G) Exon 7, (H) Exon 8, (I) to (L) Exon 9.
(TIFF)

S2 Fig. Quantitation of *ABRAXAS* mRNA levels in *ABRAXAS* knockdown MCF7 and HeLa cells. Quantitative PCR analysis of *ABRAXAS* mRNA levels in MCF7 and HeLa cells stably expressing two independent shRNAs targeting *ABRAXAS* (shABX139 and shABX145) and a control shRNA containing a non-target shRNA (shNSCTL). ACTB (beta-actin) and U6

snRNA were used as endogenous controls for normalization.
(PPTX)

S3 Fig. Quantification of gamma-H2AX formation in shABRAXAS (shABX139 and shABX145) and shNSCTL A) MCF7 and B) HeLa cells complemented with empty pOZ--HA-Flag vector after neocarzinostatin treatment and release. P-values were obtained with a Wilcoxon's Test with N = 129 cells from three independent experiments.
(TIF)

S4 Fig. Quantification of gamma-H2AX formation in shABRAXAS (shABX139 and shABX145) HeLa cells complemented with ABRAXAS-HA-Flag, ABRAXAS-HA-Flag pThr141Ile, or ABRAXAS-HA-Flag pGly39Val after neocarzinostatin treatment and release. P-values were obtained with a Wilcoxon's Test with N = 100 cells from four independent experiments.
(TIF)

S1 Table. Primers used for High Resolution Melting amplification.
(DOC)

S2 Table. Subjects excluded because of poor mutation screening performance, by study center.
(DOC)

S3 Table. ABRAXAS protein multiple sequence alignment characterization.
(DOC)

S4 Table. Primers used for subcloning exons 1, 2 and 3 into p.cDNA3.1 (+).
(DOC)

Acknowledgments

J.S. is Chairholder of the Canada Research Chair in Oncogenetics. JYM is a FRQS Chercheur National. We thank Roger Greenberg for the ABRAXAS-HA-Flag DNA construct. We wish to thank all participants in the BCFR for their contribution to the study. The BCFR would like to thank Drs. Sandra Buys—Department of Medicine, Huntsman Cancer Institute, University of Utah School of Medicine (PI—Utah site of the BCFR), Mary Daly—Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA (PI—PA site of the BCFR), and Mary Beth Terry—Department of Epidemiology, Columbia University Mailman School of Public Health, Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, USA (PI—New York site of the BCFR) and Dr. Esther John—Cancer Prevention Institute of California, Fremont, USA, Stanford University School of Medicine and Stanford Cancer Institute, Stanford, USA (contact PI for the BCFR—Esther.John@cpic.org). We thank also M. Pertesi, N. Forey, N. Robinot, G. Durand, M. Tranchant, G. Reimnitz and S. Dubois for their technical expertise, M. Leclerc for performing statistical analyses for the luciferase reporter assays and P. Belleau for help with the statistical analyses for the immunofluorescence assays.

Author Contributions

Conceived and designed the experiments: ALR YH PS JYM FL SVT JS SG. Performed the experiments: ALR YC SD. Analyzed the data: ALR PS JYM FL SVT JS. Contributed reagents/materials/analysis tools: FLCK CV MV MCS ILA EMJ JLH SG BCFR. Wrote the paper: ALR PS JYM FL SVT JS.

References

1. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet.* 2001; 27:247–254. PMID: [11242102](#)
2. Feunteun J, Lenoir GM. BRCA1, a gene involved in inherited predisposition to breast and ovarian cancer. *Biochim Biophys Acta.* 1996; 1242:177–180. PMID: [8603070](#)
3. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med.* 2008; 359:2143–2153. doi: [10.1056/NEJMra0802968](#) PMID: [19005198](#)
4. Couch FJ, Farid LM, DeShano ML, Tavtigian SV, Calzone K, Campeau L, et al. BRCA2 germline mutations in male breast cancer cases and breast cancer families. *Nat Genet.* 1996; 13:123–125. PMID: [8673091](#)
5. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature.* 1995; 378:789–792. PMID: [8524414](#)
6. Borresen AL, Andersen TI, Garber J, Barbier-Piroux N, Thorlacius S, Eyfjörd J, et al. Screening for germ line TP53 mutations in breast cancer patients. *Cancer Res.* 1992; 52:3234–3236. PMID: [1591732](#)
7. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegelé C, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* 2011; 13:R6. doi: [10.1186/bcr2810](#) PMID: [21244692](#)
8. Meijers-Heijboer H, Wijnen J, Vasen H, Wasielewski M, Wagner A, Hollestelle A, et al. The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet.* 2003; 72:1308–1314. PMID: [12690581](#)
9. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet.* 2006; 38:873–875. PMID: [16832357](#)
10. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet.* 2009; 85:427–446. doi: [10.1016/j.ajhg.2009.08.018](#) PMID: [19781682](#)
11. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet.* 2006; 38:1239–1241. PMID: [17033622](#)
12. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet.* 2007; 39:165–167. PMID: [17200668](#)
13. Tischkowitz M, Capanu M, Sabbaghian N, Li L, Liang X, Vallée MP, et al. Rare germline mutations in PALB2 and breast cancer risk: a population-based study. *Hum Mutat.* 2012; 33:674–680. doi: [10.1002/humu.22022](#) PMID: [22241545](#)
14. Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegelé C, Young EL, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res.* 2014; 16:R58. doi: [10.1186/bcr3669](#) PMID: [24894818](#)
15. Hu X, Kim JA, Castillo A, Huang M, Liu J, Wang B. NBA1/MERIT40 and BRE interaction is required for the integrity of two distinct deubiquitinating enzyme BRCC36-containing complexes. *J Biol Chem.* 2011; 286:11734–11745. doi: [10.1074/jbc.M110.200857](#) PMID: [21282113](#)
16. Mok MT, Henderson BR. The in vivo dynamic organization of BRCA1-A complex proteins at DNA damage-induced nuclear foci. *Traffic.* 2012; 13:800–814. doi: [10.1111/j.1600-0854.2012.01355.x](#) PMID: [22420687](#)
17. Wang B, Elledge SJ. Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brc1/Brcc36 complex in response to DNA damage. *Proc Natl Acad Sci U S A.* 2007; 104:20759–20763. PMID: [18077395](#)
18. Coleman KA, Greenberg RA. The BRCA1-RAP80 complex regulates DNA repair mechanism utilization by restricting end resection. *J Biol Chem.* 2011; 286:13669–13680. doi: [10.1074/jbc.M110.213728](#) PMID: [21335604](#)
19. Harris JL, Khanna KK. BRCA1 A-complex fine tunes repair functions of BRCA1. *Aging (Albany NY).* 2011; 3:461–463.
20. Hu Y, Scully R, Sobhian B, Xie A, Shestakova E, Livingston DM. RAP80-directed tuning of BRCA1 homologous recombination function at ionizing radiation-induced nuclear foci. *Genes Dev.* 2011; 25:685–700. doi: [10.1101/gad.2011011](#) PMID: [21406551](#)

21. Dever SM, Golding SE, Rosenberg E, Adams BR, Idowu MO, Quillin JM, et al. Mutations in the BRCT binding site of BRCA1 result in hyper-recombination. *Aging (Albany NY)*. 2011; 3:515–532.
22. Greenberg RA. Recognition of DNA double strand breaks by the BRCA1 tumor suppressor network. *Chromosoma*. 2008; 117:305–317. doi: [10.1007/s00412-008-0154-8](https://doi.org/10.1007/s00412-008-0154-8) PMID: [18369654](https://pubmed.ncbi.nlm.nih.gov/18369654/)
23. Kim H, Huang J, Chen J. CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. *Nat Struct Mol Biol*. 2007; 14:710–715. PMID: [17643122](https://pubmed.ncbi.nlm.nih.gov/17643122/)
24. Novak DJ, Sabbaghian N, Maillet P, Chappuis PO, Foulkes WD, Tischkowitz M. Analysis of the genes coding for the BRCA1-interacting proteins, RAP80 and Abraxas (CCDC98), in high-risk, non-BRCA1/2, multiethnic breast cancer cases. *Breast Cancer Res Treat*. 2009; 117:453–459. doi: [10.1007/s10549-008-0134-y](https://doi.org/10.1007/s10549-008-0134-y) PMID: [18695986](https://pubmed.ncbi.nlm.nih.gov/18695986/)
25. Liu Z, Wu J, Yu X. CCDC98 targets BRCA1 to DNA damage sites. *Nat Struct Mol Biol*. 2007; 14:716–720. PMID: [17643121](https://pubmed.ncbi.nlm.nih.gov/17643121/)
26. Castillo A, Paul A, Sun B, Huang TH, Wang Y, Yazinski SA, et al. The BRCA1-interacting protein Abraxas is required for genomic stability and tumor suppression. *Cell Rep*. 2014; 8:807–817. doi: [10.1016/j.celrep.2014.06.050](https://doi.org/10.1016/j.celrep.2014.06.050) PMID: [25066119](https://pubmed.ncbi.nlm.nih.gov/25066119/)
27. Solyom S, Aressy B, Pylkas K, Patterson-Fortin J, Hartikainen JM, Kallioniemi A, et al. Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions. *Sci Transl Med*. 2012; 4:122ra23. doi: [10.1126/scitranslmed.3003223](https://doi.org/10.1126/scitranslmed.3003223) PMID: [22357538](https://pubmed.ncbi.nlm.nih.gov/22357538/)
28. Pennington KP, Walsh T, Harrell MI, Lee MK, Pennil CC, Rendi MH, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res*. 2014; 20:764–775. doi: [10.1158/1078-0432.CCR-13-2287](https://doi.org/10.1158/1078-0432.CCR-13-2287) PMID: [24240112](https://pubmed.ncbi.nlm.nih.gov/24240112/)
29. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res*. 2004; 6:R375–R389. PMID: [15217505](https://pubmed.ncbi.nlm.nih.gov/15217505/)
30. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet*. 2012; 90:734–739. doi: [10.1016/j.ajhg.2012.02.027](https://doi.org/10.1016/j.ajhg.2012.02.027) PMID: [22464251](https://pubmed.ncbi.nlm.nih.gov/22464251/)
31. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. <http://evs.gs.washington.edu/EVS>
32. T-Coffee. M-Coffee a Meta-Multiple Sequence Alignment Tool. <http://www.tcoffee.org/Projects/mcoffee/index.html>.
33. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*. 2006; 34:1692–1699. PMID: [16556910](https://pubmed.ncbi.nlm.nih.gov/16556910/)
34. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB, Group IUGVW. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat*. 2008; 29:1327–1336. doi: [10.1002/humu.20892](https://doi.org/10.1002/humu.20892) PMID: [18951440](https://pubmed.ncbi.nlm.nih.gov/18951440/)
35. Grantham R. Amino acid difference formula to help explain protein evolution. *Science*. 1974; 185:862–864. PMID: [4843792](https://pubmed.ncbi.nlm.nih.gov/4843792/)
36. Ng PC. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*. 2003; 31:3812–3814. PMID: [12824425](https://pubmed.ncbi.nlm.nih.gov/12824425/)
37. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/)
38. Zhang MQ. Statistical features of human exons and their flanking regions. *Hum Mol Genet*. 1998; 7:919–932. PMID: [9536098](https://pubmed.ncbi.nlm.nih.gov/9536098/)
39. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004; 11:377–394. PMID: [15285897](https://pubmed.ncbi.nlm.nih.gov/15285897/)
40. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997; 4:311–323. PMID: [9278062](https://pubmed.ncbi.nlm.nih.gov/9278062/)
41. Perteau M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*. 2001; 29:1185–1190. PMID: [11222768](https://pubmed.ncbi.nlm.nih.gov/11222768/)
42. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009; 37:e67. doi: [10.1093/nar/gkp215](https://doi.org/10.1093/nar/gkp215) PMID: [19339519](https://pubmed.ncbi.nlm.nih.gov/19339519/)
43. Werner T. Finding and decrypting of promoters contributes to the elucidation of gene function. *In Silico Biol*. 2002; 2:249–255. PMID: [12542410](https://pubmed.ncbi.nlm.nih.gov/12542410/)

44. Pinheiro J B D, DebRoy S, Sarkar D and R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version. 2013; 3:1–111.
45. Osorio A, Barroso A, Garcia MJ, Martinez-Delgado B, Urioste M, Benitez J. Evaluation of the BRCA1 interacting genes RAP80 and CCDC98 in familial breast cancer susceptibility. *Breast Cancer Res Treat.* 2009; 113:371–376. doi: [10.1007/s10549-008-9933-4](https://doi.org/10.1007/s10549-008-9933-4) PMID: [18270812](https://pubmed.ncbi.nlm.nih.gov/18270812/)
46. Feng L, Huang J, Chen J. MERIT40 facilitates BRCA1 localization and DNA damage repair. *Genes Dev.* 2009; 23:719–728. doi: [10.1101/gad.1770609](https://doi.org/10.1101/gad.1770609) PMID: [19261748](https://pubmed.ncbi.nlm.nih.gov/19261748/)
47. Wang B, Hurov K, Hofmann K, Elledge SJ. NBA1, a new player in the Brca1 A complex, is required for DNA damage resistance and checkpoint control. *Genes Dev.* 2009; 23:729–739. doi: [10.1101/gad.1770309](https://doi.org/10.1101/gad.1770309) PMID: [19261749](https://pubmed.ncbi.nlm.nih.gov/19261749/)
48. McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G, et al. A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet.* 2011; 7:e1001333. doi: [10.1371/journal.pgen.1001333](https://doi.org/10.1371/journal.pgen.1001333) PMID: [21437268](https://pubmed.ncbi.nlm.nih.gov/21437268/)

L'Ataxie-Télangiectasie (A-T) est une maladie génétique qui s'accompagne d'un certain nombre d'anomalies neurologiques et cutanées ainsi que de désordres biologiques. L'A-T est caractérisée par le fait qu'un enfant atteint a reçu de chacun de ses parents une copie d'un gène modifié, impliqué dans l'apparition de la maladie.

Les parents d'un enfant atteint sont donc obligatoirement porteurs d'une copie du gène modifié et sont dits « hétérozygotes pour les gènes de l'A-T ». Ils ne présentent aucun symptôme de la maladie, mais auraient, comme l'ont suggéré plusieurs études, un risque augmenté de cancer (particulièrement chez les femmes).

La confirmation de ces observations pourrait avoir un retentissement important au niveau de la population générale et particulièrement lorsque l'on sera en mesure d'identifier directement les personnes hétérozygotes pour les gènes de l'A-T. En effet, bien que l'A-T soit une maladie rare, ces personnes, qui auraient elles aussi une prédisposition à l'apparition de cancers, pourraient être nombreuses, représentant jusqu'à 1% de la population générale.

Nous souhaitons pouvoir estimer plus précisément le risque de cancers, c'est pourquoi nous organisons, en concertation avec les médecins qui prennent en charge les enfants atteints d'AT, **une nouvelle étude épidémiologique de suivi de femmes appartenant à des familles A-T**, intitulée « Etude CoF-AT ».

En théorie, les progrès de la génétique permettent actuellement d'identifier les sujets porteurs d'un gène modifié. En pratique, pour les gènes de l'A-T, cette identification est réalisable uniquement dans les familles d'enfants atteints d'A-T, et pas encore dans la population générale. C'est pour cette raison que nous nous tournons vers ces familles et que leur collaboration est essentielle pour que nous puissions réaliser cette étude.

Ainsi, si vous êtes apparentée à un enfant atteint d'A-T, nous souhaitons vivement obtenir votre participation à cette recherche, qui ne pourra être menée à bien qu'avec votre collaboration.

Pour recevoir par courrier les documents qui vous donneront plus d'informations sur cette étude et vous permettront d'y participer, nous vous remercions de bien vouloir compléter le coupon-réponse ci-dessous et de nous le retourner dans l'enveloppe T jointe.

Nadine Andrieu (Inserm / Institut Curie), **Pr Jacques-Olivier Bay** (Centre J. Perrin), **Dr Nicolas Janin** (UCL), **Janet Hall** (Inserm / Institut Curie), **Pr Dominique Stoppa-Lyonnet** (Institut Curie).

*Pour tout renseignement, contactez Eve Cavaciuti : Tél. : 01 72 38 93 84 - E-mail : eve.cavaciuti@curie.fr
ou Dorothee Le Gal : Tél. : 01 72 38 93 88 - E-mail : dorothee.legal@curie.fr - Fax. : 01 53 10 26 63
Inserm U900 & Service de Biostatistiques - Institut Curie -26, rue d'Ulm, 75248 Paris cedex 05*



Coupon-Réponse à compléter et à nous retourner dans l'enveloppe T

J'accepte d'être contactée par le groupe coordonnateur de l'étude CoF-AT, mes coordonnées sont les suivantes :

Nom : Prénom :

Date de naissance :

Adresse :

Code postal : Ville :

Numéro de téléphone - domicile (ou portable) : - professionnel :

E-mail :

Nom de l'enfant atteint d'A-T : Lien avec l'enfant :

CONSENTEMENT DE PARTICIPATION

de Madame (nom, prénom) :N° famille :

Nom de jeune fille :N° sujet :

Adresse :

Le Docteur m'a proposé de participer à une recherche coordonnée par l'Institut National de la Santé Et de la Recherche Médicale (Inserm). Il (elle) m'a précisé que j'étais libre d'accepter ou de refuser. J'ai eu la possibilité de consulter la note d'information préparée à cet effet et de poser toutes les questions complémentaires auxquelles j'ai eu des réponses claires qui m'ont satisfaite.

J'ai reçu et j'ai pris connaissance des informations suivantes :

L'étude à laquelle il m'est demandé de participer a pour buts de proposer un suivi avec une action de dépistage précoce de cancer du sein (CS), d'estimer précisément le risque de cancer du sein chez les porteurs d'une mutation et d'étudier l'histoire naturelle du CS.

Il m'est demandé :

1. d'avoir un entretien avec un médecin généticien pour reconstituer mon histoire familiale,
2. de répondre à un questionnaire épidémiologique concernant mes antécédents médicaux, ma vie reproductive et mes examens radiologiques antérieurs,
3. de me faire faire une prise de sang (volume = 20 ml) si mon statut vis-à-vis de l'AT n'a pas déjà été établi,
4. d'informer ma mère, mes sœurs et nièces, mes tantes et cousines ainsi que mes filles (de 18 ans ou plus) et de leur remettre une note d'information sur l'étude. Si je suis mère d'un enfant atteint de l'AT, il m'est également demandé d'informer le père de mon enfant afin qu'il contacte les femmes de sa famille (mère, sœurs et nièces, tantes et cousines) et qu'il leur remette une note d'information sur l'étude,
5. d'être suivie pendant 10 ans, c'est-à-dire qu'une fois tous les deux ans, un questionnaire me sera adressé, mon histoire familiale sera mise à jour et il me sera pratiqué un examen clinique à partir de 25 ans et une mammographie à partir de 40 ans.

J'autorise le médecin coordonnateur à se procurer les confirmations de diagnostic concernant mes antécédents de maladie et ceux de mes ascendants décédés.

J'autorise les médecins de cette étude à contacter mon médecin généraliste ou mon gynécologue.

Je m'engage à avertir le centre coordonnateur des changements d'adresse, ainsi que des éventuelles modifications de mon état de santé survenant entre chaque suivi (exemple : survenue d'une tumeur).

J'autorise l'utilisation des prélèvements biologiques (prélèvements sanguins et éventuellement tumoraux) à des fins de recherche sur les gènes de l'AT et des autres gènes interagissant avec ceux-ci.

J'autorise également l'utilisation anonyme de l'ADN, en France et à l'étranger, sous le contrôle de l'investigateur principal, le Pr Stoppa-Lyonnet. Les recherches pourront être menées en collaboration avec des partenaires publiques ou privés et les échantillons biologiques pourront être échangés pour permettre une recherche plus rapide ou plus performante. L'ADN prélevé ne sera pas utilisé à d'autres fins que celles de la présente étude sans qu'un nouveau consentement ne me soit demandé.

Le résumé de l'entretien de génétique et les réponses aux questionnaires seront consignés dans un dossier informatisé géré par l'équipe Inserm U900 partiellement localisée dans le Service de Biostatistiques de l'Institut Curie. Ces informations seront codées de façon à les rendre anonymes. Je pourrai à tout moment vérifier l'exactitude des données et je pourrai exiger que ces informations soient supprimées.

L'étude à laquelle je participe a pour promoteur l'Inserm (qui a souscrit une assurance conformément à la loi du 20 décembre 1988) et a reçu un avis favorable du CCPPRB (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale) de Paris-Cochin en date du 26/07/2002 (Avis n° 1989), du CCTIS (Comité Consultatif sur le Traitement de l'Information en matière de recherche dans le domaine de la Santé) en date du 04/09/2002 (Avis n°02.256) et de la CNIL (Commission Nationale de l'Informatique et des Libertés) en date du 30/12/2002 (Avis n°902310). Cette étude bénéficie du soutien de l'Inserm, du Ministère de la Recherche, de la Fondation de France, du Conseil Scientifique de Radioprotection d'EDF, La Ligue Nationale Contre le Cancer, de la MGEN, de l'ITMO Santé Publique d'AVIESAN, de l'Agence Nationale de la Recherche et de l'Institut Curie.

J'ACCEPTE DE PARTICIPER A CETTE RECHERCHE DANS LES CONDITIONS PRECISEES CI-DESSUS.

Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités.

Je conserve tous mes droits garantis par la loi.

Si je le désire, je serai libre à tout moment d'arrêter ma participation. J'en informerai alors le médecin du Centre Coordonnateur de l'étude.

Les données qui me concernent resteront strictement confidentielles. Je n'autorise leur consultation que par des personnes qui collaborent à la recherche, désignées par les organisateurs, et éventuellement par un représentant des Autorités de Santé.

Je pourrai à tout moment demander toute information complémentaire aux médecins organisateurs de l'étude et demander, lorsqu'ils seront disponibles, les résultats des analyses me concernant, leur signification ainsi que les résultats de l'ensemble de l'étude.

Nom, prénom du médecin investigateur :

Signature du médecin investigateur :

Fait à..... **Le**

Signature * :

* La loi oblige tout responsable d'une recherche biomédicale à informer les participants des objectifs de cette recherche. Votre signature atteste que cette information vous a été fournie.

NOTE D'INFORMATION

L'Ataxie-Télangiectasie (AT) est une maladie qui s'accompagne d'un certain nombre de manifestations cliniques et biologiques dont une sensibilité aux radiations ionisantes (aux rayons X, par exemple). C'est une maladie génétique caractérisée par le fait qu'un enfant atteint a reçu de chacun de ses parents une copie d'un gène modifié, impliqué dans l'apparition de la maladie.

Les parents d'un enfant atteint sont donc porteurs obligatoires d'une copie du gène modifié et sont appelés hétérozygotes pour le gène AT. Ils ne présentent aucune manifestation clinique de la maladie, mais semblent avoir une sensibilité discrètement augmentée à l'action des radiations. Il apparaît également, comme l'ont suggéré plusieurs études, dont l'étude française à laquelle vous avez peut-être participé, que les hétérozygotes AT présentent un risque augmenté de cancer et tout particulièrement de cancer du sein chez les femmes.

Nous souhaitons pouvoir estimer plus précisément ce risque (qui est de l'ordre de trois par rapport celui de la population générale), c'est pourquoi nous mettons en place une étude épidémiologique de suivi de femmes appartenant à des familles AT. Ce suivi consistera en une mammographie biennale à partir de l'âge de 40 ans, ou d'un examen clinique avant cet âge. Il nous paraît raisonnable que les femmes apparentées à un enfant atteint de l'AT débutent une surveillance mammaire à partir de l'âge de 40 ans, soit un peu plus précocement que ce qui est recommandé dans la population générale. Il n'y a pas lieu de contre-indiquer les radiographies, dont les mammographies, utiles au diagnostic et aux surveillances régulières. Actuellement, aucune étude épidémiologique ne met en évidence une augmentation du risque de cancer du sein liée aux mammographies, alors que le retard au diagnostic peut avoir des conséquences très préjudiciables.

Cette recherche vise également à étudier l'histoire naturelle du cancer du sein.

Les progrès de la génétique permettent actuellement d'identifier les sujets porteurs d'un gène modifié. Pour les gènes de l'AT, cette détermination est réalisable dans les familles d'enfants malades.

La confirmation de ces données pourrait avoir un retentissement important au niveau de la population générale et particulièrement lorsque l'on sera en mesure d'identifier directement les porteurs d'un gène modifié. En effet, bien que l'AT soit une maladie rare, les sujets porteurs d'un gène AT pourraient être nombreux et représenter de 0,4 à 1 % de la population. De même que les parents d'enfants AT, ces sujets porteurs ne présentent pas de manifestations cliniques et ne se distinguent pas de la population générale mais pourraient présenter une prédisposition à l'apparition de tumeurs. Comme nous ne sommes pas, aujourd'hui, en mesure de détecter les sujets porteurs d'un gène AT dans la population générale, nous sommes amenés à nous tourner vers les familles d'enfants atteints d'AT pour entreprendre notre étude.

Indépendamment d'une prédisposition génétique aux cancers, certains facteurs ont été décrits augmentant le risque de développer une tumeur. Afin d'identifier le rôle « propre » des gènes de l'AT, il est nécessaire d'évaluer l'importance de ces facteurs pour chacun des sujets participants à l'étude. Ces facteurs (exemples : facteurs de la vie reproductive, l'exposition aux rayonnements etc...) sont évalués à l'aide d'un questionnaire adressé à chaque participante.

Vous êtes libre de participer ou non à cette étude, votre participation se faisant selon un consentement éclairé signé par vous et le médecin.

Votre participation à cette étude consiste donc :

- à avoir, avec un médecin généticien, un entretien qui retracera votre histoire familiale et au cours duquel vous bénéficierez d'un conseil génétique adapté. Cet entretien nous permettra de contacter, sous réserve de leur accord, les autres membres de votre famille (votre mère, vos sœurs et nièces, vos tantes et cousines ainsi que vos filles). Si vous êtes mère d'un enfant atteint de l'AT, il vous sera également demandé d'informer le père de votre enfant afin qu'il contacte les femmes de sa famille (mère, sœurs et nièces, tantes et cousines).

- à répondre à un questionnaire épidémiologique concernant en particulier vos antécédents médicaux, votre vie reproductive et vos examens radiologiques antérieurs.

- à vous faire faire une prise de sang (volume = 20 ml) si votre statut vis-à-vis de l'AT n'a pas déjà été établi.

- à être suivie pendant 10 ans, c'est-à-dire qu'une fois tous les deux ans, un questionnaire vous sera adressé, votre histoire familiale sera mise à jour et il vous sera pratiqué un examen clinique à partir de 25 ans et une mammographie à partir de 40 ans. En France, le dépistage du cancer du sein par mammographie est proposé habituellement aux femmes de 50 ans ou plus et son efficacité a été démontrée. Bien que chez les femmes âgées de 40 à 50 ans cette efficacité soit encore discutée, l'examen mammographique reste l'examen le plus fiable et le moins coûteux.

- à avertir le centre coordonnateur des changements d'adresse, ainsi que des éventuelles modifications de votre état de santé survenant entre chaque suivi (exemple : survenue d'une tumeur).

- à autoriser l'utilisation anonyme de l'ADN en France et à l'étranger (l'ADN prélevé ne sera pas utilisé à d'autres fins que celles de la présente étude sans qu'un nouveau consentement ne vous soit demandé).

- à autoriser les médecins de cette étude à contacter votre médecin généraliste ou votre gynécologue.

Les frais de transports et les frais annexes imputables à cette étude seront pris en charge à votre demande.

Dans le cas où la mammographie ne serait pas prise en charge en totalité par les organismes gérant votre assurance maladie et votre mutuelle, la différence vous serait remboursée à votre demande (demande à adresser au Centre Coordonnateur).

Le résumé de l'entretien de génétique et les réponses aux questionnaires seront consignés dans un dossier informatisé géré par l'équipe Inserm U900 partiellement localisée dans le Service de Biostatistiques de l'Institut Curie. Ces informations seront codées de façon à les rendre anonymes. Vous pourrez à tout moment vérifier l'exactitude des données et vous pourrez exiger que ces informations soient supprimées. Les données qui vous concernent resteront strictement confidentielles. Le fichier a été déclaré à la CNIL (Commission Nationale de l'Informatique et des Libertés), avis n°902310, le droit d'accès et de rectification qui s'applique est conforme à l'article 40 de la loi du 6 janvier 1978.

Si vous le désirez, vous serez libre à tout moment d'arrêter votre participation. Vous en informerez alors le médecin du Centre Coordonnateur de l'étude.

Vous pourrez à tout moment demander toute information complémentaire aux médecins organisateurs de l'étude et demander, lorsqu'ils seront disponibles, les résultats des analyses vous concernant, leur signification ainsi que les résultats de l'ensemble de l'étude.

Remarque : Au cours de ce projet, s'il s'avérait important d'élargir l'étude aux hommes, nous vous contacterions dans le but d'inclure les membres masculins de votre famille.

Cette étude a pour promoteur l'Inserm (qui a souscrit une assurance conformément à la loi du 20 décembre 1988) et a reçu un avis favorable du CCPPRB (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale) de Paris-Cochin en date du 26/07/2002 (Avis n°1989), du CCTIS (Comité Consultatif sur le Traitement de l'Information en matière de recherche dans le domaine de la Santé) en date du 04/09/2002 (Avis n°02.256) et de la CNIL (Commission Nationale de l'Informatique et des Libertés) en date du 30/12/2002 (Avis n°902310). Cette étude bénéficie du soutien de l'Inserm, du Ministère de la Recherche, de la Fondation de France, du Conseil Scientifique de Radioprotection d'EDF, de La Ligue Nationale Contre le Cancer, de la MGEN, de l'ITMO Santé Publique d'AVIESAN, de l'Agence Nationale de la Recherche et de l'Institut Curie.

Etude CoF-AT

Questionnaire épidémiologique EPI1

Ce questionnaire est constitué de 41 questions. Chaque page est en double exemplaire : un premier feuillet blanc (original à remplir) comportant un recto carboné et un deuxième feuillet vert (copie à conserver).

Pour remplir ce questionnaire :

1. **Avant d'écrire**, détachez la feuille cartonnée et placez-la après les deux feuillets d'une même page pour éviter de marquer les pages suivantes.
2. Utilisez un stylo à bille afin d'impressionner le deuxième feuillet.
3. Complétez les questions. Si vous ne savez pas répondre ou si vous rencontrez des difficultés, n'hésitez pas à contacter

**Mme Eve Cavaciuti (assistante de recherche)
(Inserm Emi00-06 & Service de Biostatistiques)**

**Téléphone : 01 55 43 14 64
Télécopie : 01 55 43 14 69
E-mail : eve.cavaciuti@curie.net**

**Institut Curie
26, rue d'Ulm
75248 Paris Cedex 05
France**

L'assistante de recherche vous contactera par téléphone pour confirmer les informations du questionnaire et, si besoin, les compléter.

Le temps moyen nécessaire pour compléter ce questionnaire est estimé à 35 minutes.

3) Quel est votre niveau d'études ?Pas d'études Bac à Bac+2 Certificat d'études Bac+3 à Bac+4 BEPC-CAP Bac+5 et plus

[—]

4) Quelles ont été vos différentes professions (exercées pendant un an au moins) ?

Profession (en clair)	A partir de quel âge ?	Pendant combien de temps (en années) ?	
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]
.....	[_] [_] [_]	[_] [_] [_]	[_] [_] [_] [_] [_] [_]

5) Quelle est la dernière profession de votre conjoint(e) actuel(le) ?

(avant sa retraite si il (ou elle) est retraité(e)) : [_] [_] [_] [_] [_]

6) Fumez-vous actuellement ?

Oui Non [_]

Si oui,

- . combien de cigarettes/jour ? [_] [_] [_] [_] [_] [_]
- . depuis quel âge ? [_] [_] [_] [_] [_] [_]
- . pendant combien d'années (déduire les années d'interruption)? [_] [_] [_] [_] [_] [_]

Si non,

- . avez-vous déjà fumé ? Oui Non [_]
- si oui, . combien de cigarettes/jour ? [_] [_] [_] [_] [_] [_]
- . pendant combien d'années (déduire les années d'interruption)? [_] [_] [_] [_] [_] [_]
- . à quel âge avez-vous commencé à fumer ? [_] [_] [_] [_] [_] [_]
- . depuis combien de temps avez-vous cessé de fumer (en années) ? [_] [_] [_] [_] [_] [_]

7) Quel est votre poids ? [_] [_] [_] [_] kg

[_] [_] [_] [_]

8) Quelle est votre hauteur ? [_] m [_] [_] [_] cm

[_] [_] [_] [_]

ANTECEDENTS PERSONNELS DE MALADIE

A. Réservé

9) Avez-vous eu des antécédents personnels de maladies cardio-vasculaires ?Oui Non [—]**Si oui :**

Quelle maladie ? A quel âge ?

Hypertension artérielle [—] [—] [—] [—]Hypotension artérielle [—] [—] [—] [—]Phlébite [—] [—] [—] [—]Embolie pulmonaire [—] [—] [—] [—]Infarctus du myocarde [—] [—] [—] [—]Angine de poitrine [—] [—] [—] [—]Accident vasculaire cérébral [—] [—] [—] [—]Autre (en clair) : [—] [—] [—] [—]**10) Avez-vous eu des antécédents personnels de diabète ?**Oui Non [—]**Si oui** . Quel type de diabète ? . diabète insulino-dépendant . diabète non insulino-dépendant [—]

. A partir de quel âge ? [—] [—] [—]

11) Avez-vous eu une maladie bénigne du sein ?Oui Non [—]**Si oui :**

Etait-ce ? A quel âge ?

Kyste [—] [—] [—] [—]Maladie fibrokystique [—] [—] [—] [—]Abscess au sein [—] [—] [—] [—]Mastose [—] [—] [—] [—]Autre (en clair) : [—] [—] [—] [—]Ne sais pas [—] [—] [—] [—]

- Où avez-vous été soignée ?

. Ville :

. Etablissement :

. Nom du médecin :

. Quel était votre nom usuel à cette époque ?

Confirmé

12) Avez-vous eu des antécédents personnels de polypes intestinaux?

Oui Non Ne sais pas [__]

Si oui :

Etait-ce ? *A quel âge ?*

Polypes adénomateux [__][__][__] [__][__][__]

Polypes hyperplasiques [__][__][__] [__][__][__]

Polypes vilieux [__][__][__] [__][__][__]

Ne sais pas [__][__][__] [__][__][__]

- Où avez-vous été soignée ?

. *Ville* :

. *Etablissement* :

. *Nom du médecin* :

. *Quel était votre nom usuel à cette époque ?*

Confirmé

13) Avez-vous eu un (ou des) cancer(s) ?

Oui Non [__]

Si oui (*Si plusieurs cancers, remplissez une rubrique par localisation*) :

1ère localisation (*sein, colon, poumon...*) : [__][__][__][__] . [__]

- En quelle année ? [__][__][__][__] [__][__][__][__]

- Où avez-vous été soignée ?

. *Ville* :

. *Etablissement* :

. *Nom du médecin* :

. *Quel était votre nom usuel à cette époque ?*

Confirmé

2ème localisation (*sein, colon, poumon...*) : [__][__][__][__] . [__]

- En quelle année ? [__][__][__][__] [__][__][__][__]

- Où avez-vous été soignée ?

. *Ville* :

. *Etablissement* :

. *Nom du médecin* :

. *Quel était votre nom usuel à cette époque ?*

Confirmé

B. Réserve



14) Avez-vous pris des médicaments pendant longtemps (un an ou plus) pour le traitement d'une maladie ?

Oui Non [—]

Si oui :

<i>Nom du médicament ? (en clair)</i>	<i>Age au début du traitement ?</i>	<i>Durée du traitement ? (en années)</i>	
.....	[—][—]	[—][—]	[—][—] [—][—]
.....	[—][—]	[—][—]	[—][—] [—][—]
.....	[—][—]	[—][—]	[—][—] [—][—]
.....	[—][—]	[—][—]	[—][—] [—][—]
.....	[—][—]	[—][—]	[—][—] [—][—]

VIE GENITALE

15) A quel âge avez-vous eu vos premières règles ?

[—][—]

[—][—]

. Si vous ne savez pas précisément :

- . plutôt jeune (avant 12 ans)
- . plutôt moyen (entre 12 et 14 ans)
- . plutôt tard (après 15 ans)
- . vous n'avez jamais eu de règles

. Si vous ne savez plus du tout [—]

16) Quelle est (ou était) la périodicité de vos règles (en dehors des contraceptifs oraux)?

- . Régulières avec des cycles de 25 à 31 jours
- . Régulières avec des cycles de 24 jours et moins
- . Régulières avec des cycles de 32 jours et +
- . Irrégulières
- . Ne sais plus [—]

17) Avez-vous déjà été enceinte ?

Réservé

Oui Non **Si oui,**

. Combien de fois avez-vous été enceinte ?

[__][__]

[__]

[__][__]

. Combien d'enfants nés vivants avez-vous eu ?

[__][__]

[__][__]

Remplissez le tableau ci-dessous pour chacune de vos grossesses en commençant par la première :

1 - Grossesses ayant abouti à la naissance d'un enfant (vivant ou mort-né) :

Grossesse	Age au début de la grossesse	Durée de la grossesse en mois	Date de naissance mois/an	Enfant né Vivant oui/non	Grossesse multiple notez le nombre d'enfants jumeaux = 2 triplés = 3 ...	Allaitement	
						oui/non	durée en mois
exemple	24	9	09/1970	oui	2	oui	2,5
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							

2 - Grossesses ayant abouti à une fausse couche, à une IVG (Interruption Volontaire de Grossesse) ou à une ITG (Interruption Thérapeutique de Grossesse)

Grossesse	Age au début de la grossesse	Durée de la grossesse en semaines	Date d'interruption mois/an	Interruption par une fausse couche oui/non	Interruption par une IVG ou ITG oui/non
1b					
2b					
3b					
4b					

18) Avez-vous déjà utilisé des contraceptifs oraux ?

C. Réserve

Oui Non (Si non, passez à la question 19) [__]

- A quel âge avez-vous pris des contraceptifs oraux pour la 1^{ère} fois ? [__][__] a [__][__]
ns

Si vous n'avez jamais été enceinte :

- Pendant combien de temps avez-vous pris des contraceptifs oraux (durée totale en mois) ? [__][__][__][__] [__][__][__][__]

Si vous avez été enceinte, avez-vous pris des contraceptifs oraux :

. **Avant votre 1^{ère} grossesse ?** Oui Non [__]
Si oui, pendant combien de temps (en mois) ? [__][__][__][__] [__][__][__][__]
]

. **Pendant la période située après votre 1^{ère} grossesse et avant votre dernière grossesse ?** Oui Non [__]
Si oui, pendant combien de temps (en mois) sur toute cette période ? [__][__][__][__] [__][__][__][__]
]

. **Après votre dernière grossesse ?** Oui Non [__]
Si oui, pendant combien de temps (en mois) ? [__][__][__][__] [__][__][__][__]

19) Avez-vous eu une hystérectomie (ablation de l'utérus) ?Oui Non [__]

Si oui, à quel âge ? [__][__] ans [__][__]

20) Avez-vous eu une ovariectomie (ablation d'un ou des deux ovaires) ?Oui Non [__]

Si oui : Etait-ce ? A quel âge ?

. ablation de l'ovaire gauche [__][__] [__] [__][__]

. ablation de l'ovaire droit [__][__] [__] [__][__]

. ablation des deux ovaires [__][__] [__] [__][__]

21) Etes-vous ménopausée ? (sans règles depuis 1 an ou plus)

Oui Non

[__]

Si oui, à quel âge ? [__][__] ans

[__][__]

Etait-ce ?
(plusieurs réponses possibles)

. Ménopause naturelle

[__]

. Ménopause artificielle par hystérectomie
(ablation de l'utérus)

[__]

. Ménopause artificielle par ovariectomie
(ablation des ovaires)

[__]

. Ménopause artificielle par chimiothérapie

[__]

. Ménopause artificielle par radiothérapie

[__]

. Ménopause artificielle (autre : préciser) :

[__]

22) Si vous êtes ménopausée, avez-vous pris ou prenez-vous un traitement hormonal pour la ménopause ?

Oui Non

[__]

Si oui, indiquez dans le tableau ci-dessous, les périodes successives pendant lesquelles vous avez pris un traitement hormonal de substitution, en notant l'âge de début, la durée et, si possible, le nom du médicament.

Age de début	Durée (en mois)	Nom du ou des médicament(s)
[__][__]	[__][__][__][__]
[__][__]	[__][__][__][__]
[__][__]	[__][__][__][__]
[__][__]	[__][__][__][__]
[__][__]	[__][__][__][__]
[__][__]	[__][__][__][__]

[][] [][][]

[][] [][][]

[][] [][][]

[][] [][][]

[][] [][][]

[][] [][][]

23) Coordonnées de votre médecin traitant et/ou de votre gynécologue :

Nom, Prénom du médecin traitant :

Adresse :

.....

N° de téléphone :

I.**II.** **E****t / ou :**

Nom, Prénom du gynécologue :

Adresse :

.....

N° de téléphone :

EXPOSITIONS AUX IRRADIATIONS

24) Avez-vous exercé ou exercez-vous actuellement une profession où des rayons X ou des éléments radioactifs étaient ou sont utilisés ?

Oui Non (Si non, passez à la question 25) [__]

Si oui :

. Pendant combien de temps (en années) ? [__][__] [__][__]

. A partir de quel âge ? [__][__] ans [__][__]

. Portez-vous ou avez-vous porté un badge (dosimètre individuel) ? Oui Non [__]

. Où travaillez-vous ou avez-vous travaillé ?

- Dans un hôpital, une clinique, une PMI (Protection Maternelle et Infantile) ? Oui Non [__]

- Chez un dentiste ? Oui Non [__]

- Dans un institut de recherche ? Oui Non [__]

- Dans une industrie utilisant des substances radioactives (centre nucléaire, etc) ? Oui Non [__]

Précisez en clair :

- Dans une autre structure ? Oui Non [__]

Précisez en clair :

MEDECINE PREVENTIVE OU DU TRAVAIL

25) Avez-vous subi des examens pulmonaires de type radiographique (radiographie ou scopie) dans le cadre de la médecine préventive (y compris examens pré-nuptiaux) ou dans le cadre de la médecine du travail ?

Oui Non [__]

Si oui :

. A quel âge avez-vous subi le 1^{er} examen pulmonaire ? [__][__] [__][__]

. Combien en avez-vous subi au total ? [__][__][__] [__][__][__]

. A combien de temps remonte le dernier examen (en mois) ? [__][__][__] [__][__][__]

EXAMENS RADIOGRAPHIQUES POUR PATHOLOGIES

26) Avez-vous subi des examens de type radiographique

(radiographie, scopie, tomographie, bronchographie, scintigraphie, angiographie, scanner ou autre) **pour des pathologies**

pulmonaires (tuberculose, primo-infection, corps étrangers dans l'appareil respiratoire, dilatation des bronches, bronchite chronique, emphysème, fibrose pulmonaire ou autre) ?

Oui Non Ne sais pas

[__]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]

27) Avez-vous subi des examens de type radiographique

(scintigraphie osseuse, myélographie, radiographie, tomographie, scanner ou autre) **aux côtes ou à la colonne vertébrale**

(fractures, scolioses...)?

Oui Non Ne sais pas

[__]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[__]	[__]	[__]
------	-------	-------	------	------

28) Avez-vous subi des examens de type radiographique (radiographie, arthrographie, tomographie, scanner ou autre) **au crâne, aux dents** (en dehors d'examen panoramique), **aux bras, aux jambes, au bassin** (traumatismes, fractures, arthroses, caries...) ?

Oui Non Ne sais pas

[__]

A quel âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]
[__]	[__]	[__]

29) Avez-vous subi des examens panoramiques aux dents ?

Oui Non Ne sais pas

[__]

A quel âge ?

Pour quelle raison ?
(en clair)

[__]	[__]
[__]	[__]
[__]	[__]

[]	[]
[]	.	[]
[]	[]
	.	

30) Avez-vous subi des examens de type radiographique

(scintigraphie, scanner) **pour une maladie de la thyroïde** (hypo ou hyper-thyroïdie) ?

Oui Non Ne sais pas [—]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[—] [—] [—]

[—] [—] [—]

[—] [—] [—]

. Si vous avez subi un examen radiographique de la thyroïde, cet examen a été effectué dans quel établissement ?

1^{er} examen :

. Ville, département :

. Etablissement :

. En quelle année ? [—] [—] [—] [—]

2^{ème} examen :

. Ville, département :

. Etablissement :

. En quelle année ? [—] [—] [—] [—]

3^{ème} examen :

. Ville, département :

. Etablissement :

. En quelle année ? [—] [—] [—] [—]



31) Avez-vous subi des examens de type radiographique

(radiographie, scopie, scintigraphie, coronarographie, artériographie, phlébographie ou autre) **au cœur et aux vaisseaux** (infarctus, phlébite, embolie pulmonaire, malformation cardiaque, souffle au cœur, etc.) ?

Oui Non Ne sais pas

[—]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

32) Avez-vous subi des examens de type radiographique

(radiographie, transit du grêle, lavement baryté, TOGD (Transit Oeso-GastroDuodéнал) ou autre) **aux intestins et à l'estomac** (ulcère, diverticules, polypes, maladies inflammatoires ou autre) ?

Oui Non Ne sais pas

[—]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[—][—]

.....

[—][—]

[—][—]

.

[—][—]

.....

[—][—]

[—][—]

.

33) Avez-vous subi des examens de type radiographique
(radiographie, cholécystographie ou autre) **à la vésicule biliaire,**
foie, pancréas (calculs biliaires, jaunisse, cholécystite ou autre) ?

Oui Non Ne sais pas

[—]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

34) Avez-vous subi des examens de type radiographique

(urographie, artériographie rénale, scanner ou autre) **pour une**
maladie des reins, de la vessie, du système urinaire

(traumatisme rénal, malformation urogénitale, infection urinaire,
colique néphrétique ou autre) ?

Oui Non Ne sais pas

[—]

A quel
âge ?

Quel type d'examen ?
(en clair)

Pour quelle raison ?
(en clair)

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

.

[—][—][—]

.....

[—][—][—]

[—][—][—]

35) Avez-vous subi des examens de type radiographique (hystéroggraphie ou autre) à l'utérus (en dehors des grossesses), aux appareils génitaux ?

Oui Non Ne sais pas [—]

<i>A quel âge ?</i>	<i>Quel type d'examen ? (en clair)</i>	<i>Pour quelle raison ? (en clair)</i>	[—]	[—]
[—][—][—]	[—][—][—]	[—][—][—]
[—][—][—]	[—][—][—]	[—][—][—]
[—][—][—]	[—][—][—]	[—][—][—]

36) Avez-vous subi des examens de type radiographique à l'utérus pendant une grossesse (pelvimétrie) ?

Oui Non Ne sais pas [—]

<i>A quel âge ?</i>	<i>Combien d'examens ?</i>	<i>Pour quelle grossesse ? (préciser laquelle : cf page 6)</i>	[—]	[—]	[—]
[—][—][—]	[—][—][—]	[—][—][—]	[—]	[—]	[—]
[—][—][—]	[—][—][—]	[—][—][—]	[—]	[—]	[—]
[—][—][—]	[—][—][—]	[—][—][—]	[—]	[—]	[—]

37) Avez-vous subi des examens de type radiographique pour des maladies non déjà citées, en dehors des mammographies ?

Oui Non Ne sais pas [—]

<i>A quel âge ?</i>	<i>Quel type d'examen ? (en clair)</i>	<i>Pour quelle raison ? (en clair)</i>	[—]	[—]
[—][—][—]	[—][—][—]	[—][—][—]
[—][—][—]	[—][—][—]	[—][—][—]
[—][—][—]	[—][—][—]	[—][—][—]

[]	[]	[]
	.			
[]	[]	[]
	.			
[]	[]	[]
	.			

EXAMENS RADIOGRAPHIQUES AUX SEINS

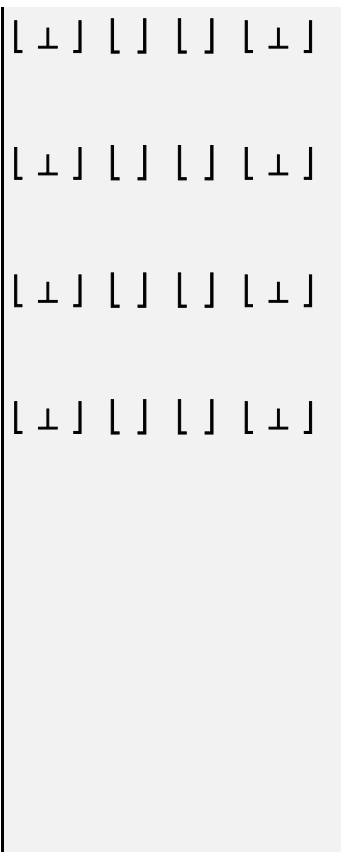
38) Avez-vous subi des mammographies (radio des seins) ?

Oui Non Ne sais pas [—]

Si oui : précisez à quel âge, à quel sein (droit = D, gauche = G, les deux = DG), pour quelle raison et si vous le pouvez, combien de clichés étaient pris par examen :

	Age	Sein	Raison de la mammographie	Nombre de clichés	
1 ^{ère} mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
2 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
3 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
4 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
5 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
6 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
7 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
8 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
9 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
10 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		
11 ^e mammo	[—][—]	[—][—]	[—][—]	[] [] [] [] []
]]		

12 ^e mammo	[_]	[_]	[_]
			
13 ^e mammo	[_]	[_]	[_]
			
14 ^e mammo	[_]	[_]	[_]
			
15 ^e mammo	[_]	[_]	[_]
			



TRAITEMENTS PAR RAYONNEMENTS IONISANTS
(radiothérapie, curiethérapie, iode radioactif...)

39) Avez-vous déjà été soignée par des rayons pour une maladie bénigne (angiomes...) ?

Oui Non Ne sais pas [—]

A quel âge ? Raison du traitement ?
(en clair)

Tête (cerveau, maxillaire, végétations...)

[—] [—] [—]

.....

Cou (thyroïde, pharynx...)

[—] [—] [—]

.....

Poumons

[—] [—] [—]

.....

Sein droit

[—] [—] [—]

.....

Sein gauche

[—] [—] [—]

.....

Ovaires

[—] [—] [—]

.....

Utérus

[—] [—] [—]

.....

Peau

[—] [—] [—]

.....

Autres (précisez en clair)

[—] [—] [—]

.....

.....

....

40) Avez-vous déjà été soignée par des rayons pour une maladie maligne ?

Oui Non Ne sais pas [—]

A quel âge ? Raison du traitement ?
(en clair)

Tête (cerveau, maxillaire, végétations...)

[—] [—] [—]

.....

Cou (thyroïde, pharynx...)

[—] [—] [—]

				
Poumons	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Sein droit	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Sein gauche	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Ovaires	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Utérus	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Peau	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
				
Autres (précisez en clair)	<input type="checkbox"/>	[—┘—]	[—]	[—┘—]
.....				
....					

Annexe 6 : Nouvelles questions ajoutées aux questionnaires épidémiologiques de suivi numéro 1, 2, 3, 4 et 5.

EPI2 - 6) Quelle a été votre consommation d'alcool au cours de votre vie ?

Indiquez dans le tableau ci-dessous, pour chaque période de votre vie, la fréquence moyenne de votre consommation (en nombre de fois par semaine; par exemple : tous les soirs de la semaine = 7 fois ; à tous les repas =14 fois ; etc.) et la quantité moyenne par semaine (en nombre de verres consommés) :

	Fréquence (nombre de fois par semaine)	Quantité (nombre de verres par semaine)	
Avant l'âge de 18 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
De 18 à 30 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
De 31 à 40 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
De 41 à 50 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
De 51 à 60 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
De 61 à 70 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]
Après 70 ans :	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]	[_] [_] [_] [_] [_]

Epi2 - 8) Êtes-vous droitère ou gauchère ?

- . Droitère
- . Gauchère
- . Gauchère contrariée

[_]

EPI2 - **13) Au cours de votre vie, avez-vous eu une maladie bénigne de la thyroïde ou une intervention chirurgicale à la thyroïde?**

Oui Non []

Si oui, précisez (plusieurs réponses possibles) :

A quel âge ?

- | | | | | |
|-------------------------------------|------------------------------|------------------------------|-------------|-----------------|
| . Hyperthyroïdie | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Hypothyroïdie | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Adénome | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Nodule | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Kyste | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Goitre | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Maladie de Basedow | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Maladie de Hashimoto | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Ablation partielle de la thyroïde | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Ablation totale de la thyroïde | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Autre (précisez) : | Oui <input type="checkbox"/> | Non <input type="checkbox"/> | [] [] [] | [] [] [] [] |
| . Ne sais pas | <input type="checkbox"/> | | [] [] [] | [] [] [] [] |

. Où avez-vous été suivie ?

- . Ville, département :
- . Etablissement :
- . Nom du médecin :
- . En quelle année ? [] [] [] []

Epi3 - **3) À quel groupe de population appartenez-vous * ?**

- Caucasienne / Blanche
- Ashkénaze
- Noire-africaine
- Afro-caribéenne
- Arabe
- Berbère
- Asiatique
- Indienne / Pakistanaise

[] [] []

* Si vous êtes métisse (née d'une mère et d'un père d'origine différente), veuillez cocher les groupes dont vous êtes issue.

Epi3 - 7) Combien de vols aériens d'une durée de 4 heures et plus avez-vous effectués à chaque période de votre vie ?

Un vol comprend 1 décollage et 1 atterrissage. Si aucun vol, notez 0. Un aller-retour compte comme 2 vols. Un aller avec une escale compte comme 2 vols si chaque vol dure 4 heures ou plus, etc.

	Nombre de vols aériens (professionnels, loisirs...)	
- Avant l'âge de 18 ans :	[] [] [] []	[] [] [] []
- De 18 à 30 ans :	[] [] [] []	[] [] [] []
- De 31 à 40 ans :	[] [] [] []	[] [] [] []
- De 41 à 50 ans :	[] [] [] []	[] [] [] []
- De 51 à 60 ans :	[] [] [] []	[] [] [] []
- De 61 à 70 ans :	[] [] [] []	[] [] [] []
- Après 70 ans :	[] [] [] []	[] [] [] []

EPI3 - 27) Au cours de votre vie, avez-vous pris ou prenez-vous des phytoestrogènes (oestrogènes végétaux) pour la ménopause ? (Phytosoya®, Ymea®, Sojyam®, Evestrel®, etc.)

Oui Non

[]

Si oui, précisez :

Âge au début du traitement	Nom du ou des médicament(s)	Durée (en mois)
[] [] ans	[] [] [] []
[] [] ans	[] [] [] []
[] [] ans	[] [] [] []
[] [] ans	[] [] [] []
[] [] ans	[] [] [] []

EPI4 - **22) Votre mère a-t-elle pris du Distilbène®** (traitement prescrit pour éviter les accouchements prématurés) **pendant qu'elle était enceinte de vous ?**

Autres noms du Distilbène® : diéthylstilboestrol, Stilboestrol-Borne®

Oui Non Ne sais pas []

EPI5 - **10) Quels étaient votre poids et votre taille (hauteur) à l'âge où vous avez eu vos premières règles ?**

. Poids : [] kg Ne sais pas

. Taille : [] m [] cm Ne sais pas

EPI5 - **18) Au cours de votre vie, avez-vous eu une maladie des yeux ?**

Oui Non []

Si oui, précisez (plusieurs réponses possibles) :

Âge au diagnostic ?

. Cataracte : . œil droit Oui Non [] [] []

. œil gauche Oui Non [] [] []

. bilatérale Oui Non [] [] []

Si oui, de quel type ? (cochez la case correspondante) :

. sous-capsulaire corticale nucléaire

. autre (précisez) :

. inconnu

. Presbytie (difficulté à voir de près) Oui Non [] [] []

. Glaucome chronique Oui Non [] [] []

. Dégénérescence maculaire (D.M.L.A) Oui Non [] [] []

. Autre baisse visuelle avec déformation des lignes droites Oui Non [] [] []

Précisez en clair :

. Autre maladie affectant la rétine Précisez laquelle :	Oui <input type="checkbox"/> Non <input type="checkbox"/>	[][]	[][][][]
. Autre (précisez) :	Oui <input type="checkbox"/> Non <input type="checkbox"/>	[][]	[][][][]

De plus, la question sur la contraception a été reformulée à partir de EPI2 afin de prendre en compte les contraceptions de type hormonal autres que la pilule, telles que l'implant sous-cutané, le stérilet MIRENA, le patch etc. (non demandés à EPI1) :

EPI2 - 19) Avez-vous utilisé depuis votre dernier questionnaire ou utilisez-vous actuellement des contraceptifs hormonaux (pilule, patch, implant sous-cutané...)?			[][]
Oui <input type="checkbox"/> Non <input type="checkbox"/>			
Si oui , indiquez dans le tableau ci-dessous les périodes pendant lesquelles vous avez pris des contraceptifs hormonaux, en notant la date de début, la durée et le nom du contraceptif oral (pilule) ou non oral (exemples : bracelet ou patch transdermique, implant sous-cutané...) :			
Date de début (mois/année)	Durée (en mois)	Nom du (ou des) contraceptif(s)	
..... /	[][] / [][]
..... /	[][] / [][]
..... /	[][] / [][]
..... /	[][] / [][]

Annexe 7 : Protocole SureSelectQXT Target Enrichement complet pour la préparation des banques d'ADN

1. Préparation des échantillons

Réactifs	Température de stockage	Mode de décongélation
SureSelect QXT buffer	-20°C	sur glace
SureSelect QXT Enzyme Mix ILM	-20°C	sur glace
DMSO (utilisé à l'étape 1.4)	-20°C	TA
Bille Agencourt	+4°C	TA 30 minutes avant utilisation

1.1. Fragmentation de l'ADNg et ajout des adaptateurs

- Pour chaque échantillon mettre dans un tube PCR et en respectant l'ordre :
 - 17 μ L *SureSelect QXT Buffer*
 - 2 μ L ADNg à 25ng/ μ L à placer au fond du tube.
 - 2 μ L *SureSelect QXT Enzyme Mix ILM* (faire 8 allers-retours à la pipette pour s'assurer que la totalité de l'enzyme ait été transférée dans le tube)
- Vortexer pendant 20 s à vitesse maximale et centrifuger.
- Programme de fragmentation de l'ADN :

Etapes	Température	Temps
1	45°C	10 minutes
2	4°C	1 minute
3	4°C	∞

- Sur glace, ajouter 32 μ L *SureSelect QXTT Stop Solution* (contenant 25% d'éthanol) dans chaque tube.
- Vortexer à vitesse maximale pendant 5 secondes et centrifuger.
- Incuber les échantillons à TA pendant 1 min avant de poursuivre.

1.2. Purification des banques

- Ajouter 52 μ L de billes Agencourt par puits (contenant 53 μ L de l'ADNg fragmenté)
- Homogénéiser à la pipette par aspiration-refoulement
- Incuber 5 min à température ambiante
- Mettre les tubes sur la plaque magnétique, laisser 3 min
- Eliminer le surnageant en laissant les tubes sur la plaque magnétique
- Ajouter 200 μ L d'éthanol à 70°, laisser environ 1 min
- Eliminer le surnageant

- Refaire les étapes 6 et 7
- Fermer les tubes et faire une mini-centrifugation pour faire tomber le liquide restant. Redisposer les barrettes sur la plaque magnétique et éliminer le liquide restant à l'aide de la pipette multicanaux de 10 µL.
- Placer les barrettes sans les capuchons sur un thermocycleur régler à 37°C, laisser environ 3 min avec le couvercle ouvert
- Resuspendre les billes avec 12 µL d'eau ultra-pure. Bien homogénéiser à la pipette par aspiration-refoulement (la viscosité du *SureSelect QXT Buffer* complique cette étape).
- Laisser 2 min à TA.
- Remettre sur la plaque magnétique et laisser 2 min
- Prélever le surnageant et le transférer dans une nouvelle barrette.
- Placer les barrettes sur la glace

1.3. Amplification des banques (pièce pré-PCR)

Réactifs	Température de stockage	Mode de décongélation
Herculase II Fusion DNA Polymerase	-20°C	sur glace
5x Herculase II Reaction Buffer	-20°C	sur glace
100 mM dNTP (25 mM each dNTP)	-20°C	sur glace
SureSelect QXT Primer Mix	-20°C	sur glace

- Préparation du mélange réactionnel

Réactifs	X 1	X 16	X 32
Nuclease-free water	25 µL	425 µL	850 µL
5x Herculase II Reaction Buffer	10 µL	170 µL	340 µL
100 mM dNTP Mix	0,5 µL	8,5 µL	17 µL
DMSO	2,5 µL	42,5 µL	85 µL
SureSelect QXT Primer Mix	1 µL	17 µL	34 µL
Herculase II Fusion DNA Polymerase	1 µL	17 µL	34 µL
Total	40 µL	680 µL	1360 µL

- Distribuer 40 µL du MIX pour chaque banque purifié (10µL) à l'étape **1.3**
- Vortexer pendant 5 secondes à vitesse moyenne

- Programme d'amplification des banques

Etapes	Témpérature	Temps	
1	68°C	2 minutes	
2	98°C	2 minutes	
3	98°C	30 secondes	X 8 cycles
	57°C	30 secondes	
	72°C	1 minute	
4	72°C	5 minutes	
5	4°C	∞	

1.4. Purification des banques amplifiées

- Ajouter 50 µL de billes Agencourt par puits (contenant 50µL de la PCR précédente)
- Homogénéiser à la pipette par aspiration-refoulement
- Incuber 5 min à température ambiante
- Mettre les tubes sur la plaque magnétique, laisser 3 min
- Eliminer le surnageant en laissant les tubes sur la plaque magnétique
- Ajouter 200 µL d'éthanol à 70°, laisser environ 1 min
- Eliminer le surnageant
- Refaire les [étapes 6 et 7](#)
- Fermer les tubes et faire une mini-centrifugation pour faire tomber le liquide restant. Redisposer les barrettes sur la plaque magnétique et éliminer le liquide restant à l'aide de la pipette multicanaux de 10 µL.
- Placer les barrettes sans les capuchons sur un thermocycleur régler à 37°C, laisser environ 3 min avec le couvercle ouvert
- Resuspendre les billes avec 14 µL d'eau ultra-pure. Bien homogénéiser à la pipette par aspiration-refoulement
- Laisser 2 min à TA.
- Remettre sur la plaque magnétique et laisser 2 min
- Prélever le surnageant (régler la pipette à 13 µL) et le transférer dans une nouvelle barrette.

Les échantillons peuvent être stockés à +4°C (stockage court) ou à -20°C (stockage long) à cette étape.

1.5. Contrôles et dosages des banques

Jusqu'à 11 banques d'ADN à contrôler, utiliser une puce Bioanalyzer HS
 A partir de 12 banques d'ADN à contrôler, utiliser une puce LabChip HS.

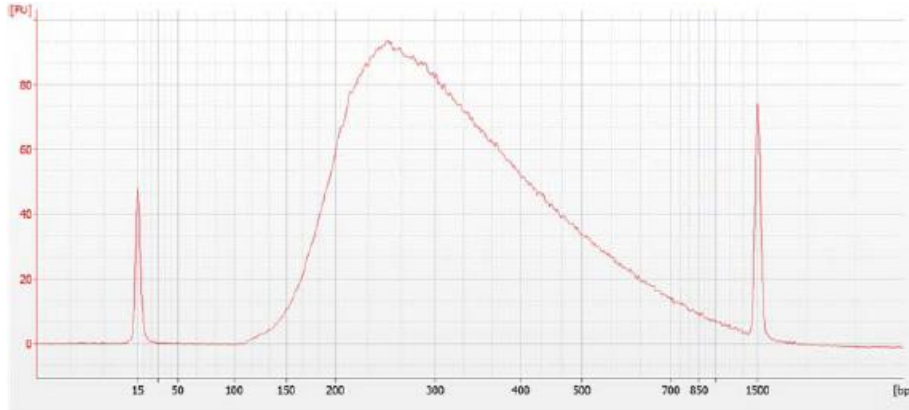
Le dosage se fait sur les échantillons dilués au 1/50^e

1.5.1. Contrôle sur puce Bioanalyzer (Agilent)

Diluer les échantillons au 1/50^e

Utiliser une puce High Sensitivity DNA

Profil attendu:

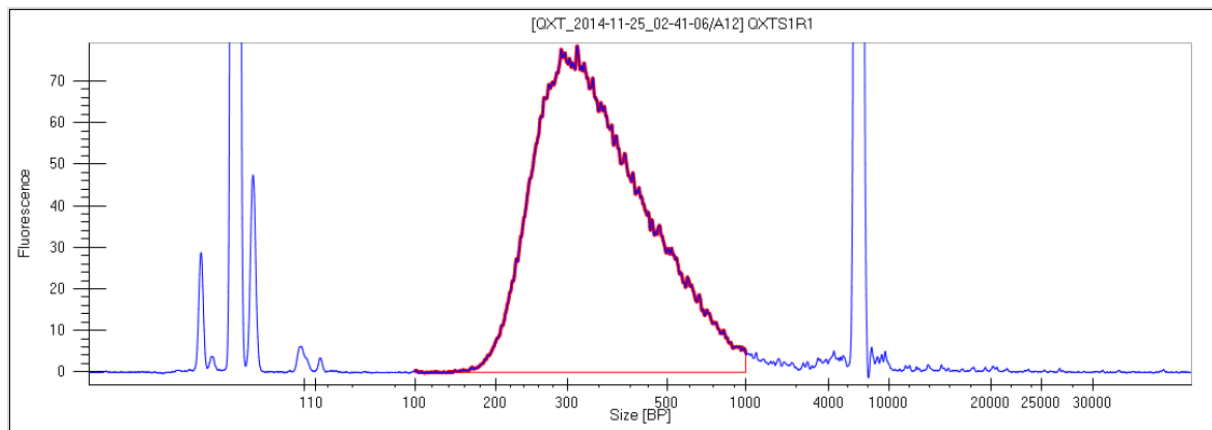


1.5.2. Contrôle sur puce Labchip (Caliper)

Diluer les échantillons au 1/50^e

Utiliser la puce HT DNA 1K/12K/HiSens

Profil attendu



2. Hybridation et capture (pièce post-PCR)

2.1. Préparation des aliquots pour l'hybridation

Dans des barrettes, mettre 650 ng d'ADN dans un volume final de 12 µL. Ajuster le volume avec de l'eau ultra-pure.

2.2. Hybridation des ADN

Réactifs	Température de stockage	Mode de décongélation
SureSelect QXT Fast Hybridization buffer	-20°C	à température ambiante
SureSelect QXT Fast Blocker Mix	-20°C	sur glace
SureSelect RNase Block	-20°C	sur glace
SureSelect ^{XT} Custom Capture Library	-80°C	sur glace

- Vortexer chaque réactifs une fois décongelé et les centrifuger.
- Ajouter 5 µL de SureSelect QXT Fast Blocker Mix dans chaque tube
- Vortexer 5 secondes à vitesse maximale et centrifuger. Volume final = 17µL

- Programme d'hybridation des ADN

Étapes	Température	Temps	
1	95°C	5 minutes	
2	65°C	10 minutes	
3	65°C	1 minute (mettre en PAUSE ici)	
4	65°C	1 minute	X 60 cycles
	37°C	3 secondes	
5	65°C	∞	

- Préparer 25% de SureSelect RNase Block selon les proportions suivantes en fonction du nombre d'ADN :

Réactif	X1	X16	X32
Nuclease-free water	1,5 µL	30 µL	60 µL
SureSelect RNase Block	0,5 µL	10 µL	20 µL
Total	2 µL	40 µL	80 µL

Réactif	X1	X16	X32
Nuclease-free water	3 µL	51 µL	102 µL
25% SureSelect RNase Block	2 µL	34 µL	68 µL
SureSelect ^{XT} Custom Capture Library	2 µL	34 µL	68 µL
SureSelect QXT Fast Hybridization buffer	6 µL	102 µL	204 µL
Total	13 µL	221 µL	442 µL

- Mettre le thermocycleur sur pause une fois le programme arrivé à l'étape 3.
- Laisser les tubes en place sur le thermocycleur à 65°C et distribuer 13 µL du Mix Capture Library Hybridization
- Vortexer les tubes, centrifuger et les remettre sur le thermocycleur.
- Relancer le programme.

2.3. Activation des billes streptavidine

- Vortexer le flacon Dynabeads MyOne Streptavidin T1 contenant les billes magnétiques.

- Distribuer 50 μL de billes par puits nécessaire
- Laver les billes (activation) :
 - Ajouter 200 μL de SureSelect Binding Buffer.
 - Mélanger par aspiration-refoulement 10 fois à l'aide d'une pipette.
 - Mettre la plaque ou les barrettes sur la plaque magnétique.
 - Attendre 5 min puis retirer le surnageant.
 - Répéter les étapes 2 fois (3 lavages auront été réalisés au total).
- Resuspendre les billes dans 200 μL de SureSelect Binding Buffer.

2.4. Capture des ADN hybridés

- Récupérer les ADN hybridés et faire la suite de la manip à TA.
- Transférer la totalité du volume ADN
- Mettre les barrettes sous agitation à 1800 rpm pendant 30 min à TA.
- Mettre de la solution **SureSelect Wash Buffer 2** à 65°C :
 - Aliquoter 3 puits par échantillon avec 200 μL de SureSelect Fast Wash Buffer 2
 - Fermer les barrettes, les disposer sur un thermocycleur et lancer un programme à 65°C en continu.
- Une fois les 30 min d'agitation terminé, faire tomber le liquide au fond du tube : par un mouvement rapide du bras s'il s'agit d'une plaque, à l'aide d'une mini-centrifugeuse pour les barrettes.
- Placer les tubes sur la plaque magnétique, laisser 1 min
- Eliminer le surnageant en laissant les barrettes sur la plaque magnétique
- Resuspendre les billes avec 200 μL de **SureSelect Wash Buffer 1** par aspiration-refoulement
- Vortexer 8 secondes et faire tomber le liquide au fond des tubes.
- Placer les tubes sur la plaque magnétique, laisser 1 min et éliminer le surnageant.
- Procéder ensuite au lavage des billes de la manière suivante :
 - Resuspendre les billes avec 200 μL de **SureSelect Wash Buffer 2** préchauffé à 65°C. Homogénéiser par aspiration-refoulement.
 - Fermer avec de nouvelles barrettes de capuchons. Vortexer 5 secondes et faire tomber le liquide au fond des tubes.
 - Incuber les échantillons 10 min à 65°C sur un thermocycleur. Fermer le couvercle du thermocycleur.
 - Mettre ensuite les tubes sur une plaque magnétique. Attendre 1 min et éliminer le surnageant.
 - Répéter les étapes a à d 2 autres fois (3 lavages au total)
- Après élimination du surnageant, centrifuger rapidement les échantillons, replacer les barrettes sur la plaque magnétique et éliminer le maximum de solution restante.
- Ajouter 23 μL d'eau ultra-pure dans chaque tube. Placer les barrettes sur la glace jusqu'à la réalisation de la PCR.

3. Indexage et préparation des banques au séquençage

3.1. Amplification des ADN capturés

3.1.1. Manip

Réactifs	Température de stockage	Mode de décongélation
Herculase II Fusion DNA Polymerase	-20°C	sur glace
5x Herculase II Reaction Buffer	-20°C	sur glace
100 mM dNTP (25 mM each dNTP)	-20°C	sur glace
SureSelect QXT P7 et P5 dual indexing primers	-20°C	sur glace

- Préparation du mélange réactionnel :

Réactifs	X 1	X 16	X 32
Nuclease-free water	13,5 µL	229,5 µL	459 µL
5x Herculase II Reaction Buffer	10 µL	170 µL	340 µL
100 mM dNTP Mix	0,5 µL	8,5 µL	17 µL
Herculase II Fusion DNA Polymerase	1 µL	17 µL	34 µL
Total	25 µL	680 µL	1360 µL

- Récupérer les barrettes contenant les ADN capturés et laisser sur glace. Distribuer 25 µL du
- MIX pour chaque échantillon.
- Ajouter 1 µL du primer P7 correspondant au barcode de l'échantillon
- Ajouter 1 µL du primer P5 correspondant au barcode de l'échantillon
- Homogénéiser à la pipette par aspiration-refoulement.

- Programme d'amplification :

Étapes	Température	Temps	
1	98°C	2 minutes	
2	98°C	30 secondes	X 14 cycles
	58°C	30 secondes	
	72°C	1 minute	
3	72°C	5 minutes	
4	4°C	∞	

a. Purification des ADN amplifiés

- Ajouter 60 µL de billes Agencourt par puits (contenant 50µL de la PCR précédente) et homogénéiser, Incuber 5 min à température ambiante
- Mettre les tubes sur la plaque magnétique. Laisser 3 min.
- Eliminer le surnageant en laissant les tubes sur la plaque magnétique
- Toujours en laissant les tubes sur la plaque magnétique, ajouter 200 µL d'éthanol à 70°. Laisser environ 1 min
- Eliminer le surnageant
- Refaire les étapes 4 et 5
- Placer les barrettes sans les capuchons sur un thermocycleur régler à 37°C, laisser environ 3 min avec le couvercle ouvert.

- Resuspendre les billes avec 26,5 μ L d'eau ultra-pure. Bien homogénéiser à la pipette.
- Centrifuger puis remettre sur la plaque magnétique, laisser 2 min
- Prélever le surnageant et le transférer dans une nouvelle barrette. Jeter les billes.

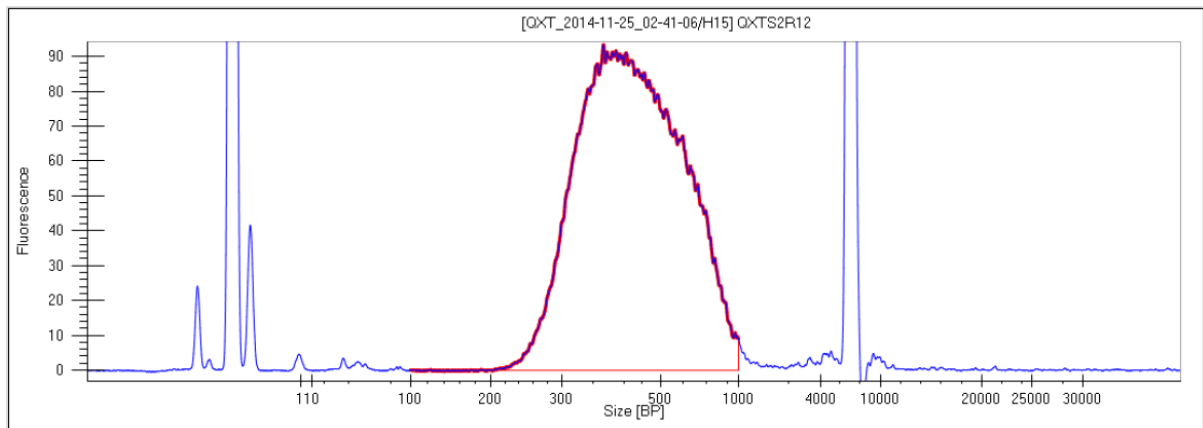
b. Quantification des banques ADN indexés par Qubit

c. Détermination de la taille des banques ADN indexés par Labchip

Diluer les échantillons au $1/10^6$

Utiliser la puce HT DNA 1K/12K/HiSens

Profil attendu :



Annexe 8 : Noms et séquences des oligonucléotides («index ») utilisés pour identifier les échantillons pour le séquençage du panel de 55 gènes.

Index	Séquence
P7 Index 1 (P7 i1)	TAAGGCGAA
P7 Index 2 (P7 i2)	CGTACTAG
P7 Index 3 (P7 i3)	AGGCAGAA
P7 Index 4 (P7 i4)	TCCTGAGC
P7 Index 5 (P7 i5)	GTAGAGGA
P7 Index 6 (P7 i6)	TAGGCATG
P7 Index 7 (P7 i7)	CTCTCTAC
P7 Index 8 (P7 i8)	CAGAGAGG
P7 Index 9 (P7 i9)	GCTACGCT
P7 Index 10 (P7 i10)	CGAGGCTG
P7 Index 11 (P7 i11)	AAGAGGCA
P7 Index 12 (P7 i12)	GGACTCCT

Index	Séquence
P5 Index 13 (P5 i13)	TAGATCGC
P5 Index 14 (P5 i14)	CTCTCTAT
P5 Index 15 (P5 i15)	TATCCTCT
P5 Index 16 (P5 i16)	AGAGTAGA
P5 Index 17 (P5 i17)	GTAAGGAG
P5 Index 18 (P5 i18)	ACTGCATA
P5 Index 19 (P5 i19)	AAGGAGTA
P5 Index 20 (P5 i20)	CTAAGCCT

Annexe 9 : Protocole de q-PCR utilisé pour la mesure de la longueur relative des télomères.

Tableau 1 : Séquences et caractéristiques des amorces utilisées pour l'amplification des télomères et du gène de l'albumine utilisé comme gène de référence.

Amorces	Séquences	Taille (pb)	Température de fusion (Tm)	Longueur amplicon (pb)
telg (F)	ACACTAAGGTTTGGGTTTGGGTTTGGGTTTGGGTTAGTGT	40	72	indéfini
telc (R)	TGTTAGGTATCCCTATCCCTATCCCTATCCCTATCCCTAACA	42	72	
albu (F)	CGGCGGCGGGCGGCGGGCTGGGCGGaaatgctgcacagaatccttg	48	68	98bp (44+2x27)
albd (R)	GCCCGGCCCGCCGCGCCCGTCCCGCCGgaaaagcatggctgcctgtt	47	67	

Tableau 2 : Mélange réactionnel GoTaq Master Mix pour l'amplification des télomères

Amplification des télomères	Concentration finale	Volume pour 1 réaction (uL)
Nuclease-Free water		2.55
qPCR Master Mix (2x)	1x	5.00
telc (20uM)	0.7uM	0.35
telg (20uM)	0.2uM	0.10
ADN (5ng/uL)		2.00
Volume		10

Tableau 3 : Mélange réactionnel GoTaq Master Mix pour l'amplification des télomères

Amplification albumine	Concentration finale	Volume pour 1 réaction (uL)
Nuclease-Free water		2.60
qPCR Master Mix (2x)	1x	5.00
AlbU (20uM)	0.3uM	0.15
AlbG (20uM)	0.5uM	0.25
ADN (5ng/uL)		2.00
Volume		10

Tableau 4 : Programme d'amplification de la q-PCR

Nombre de cycles	Temperature	Temps
1	95°	5min
40	95°	30sec
	54°	30sec
	72°	20sec
infini	40°	

Annexe 10 : Protocole de géotypage par la méthode TaqMan

Tableau 1 : Mélange réactionnel pour le géotypage par la méthode TaqMan

Nombre de réactions	Volume pour 1 réaction (uL)
2x Master Mix TaqMan	2.5
40x Genotyping Assay	0.0625
Nuclease-free Water	2.4375
Dried DNA	10ng
V total	5

Tableau 2 : Programme d'amplification pour le géotypage