HAL
open science

Analyse de l'impact sur l'épissage de l'ARN de variants de signification inconnue de gènes impliqués dans la prédisposition héréditaire aux cancers du sein et/ou de l'ovaire

Grégoire Davy

To cite this version:

Grégoire Davy. Analyse de l'impact sur l'épissage de l'ARN de variants de signification inconnue de gènes impliqués dans la prédisposition héréditaire aux cancers du sein et/ou de l'ovaire. Sciences pharmaceutiques. 2017. dumas-01557299

HAL Id: dumas-01557299

https://dumas.ccsd.cnrs.fr/dumas-01557299v1

Submitted on 6 Jul 2017

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

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

UNIVERSITE DE CAEN NORMANDIE
ANNEE 2017

U.F.R. DES SCIENCES
PHARMACEUTIQUES

MEMOIRE
DU DIPLOME D'ETUDES SPECIALISEES
D'INNOVATION PHARMACEUTIQUE ET RECHERCHE

soutenu devant le Jury Interrégional
le vingt janvier deux mille dix-sept
par Grégoire DAVY

conformément aux dispositions de l'arrêté
du 4 octobre 1988 tient lieu de

THESE POUR LE DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE

Analyse de l'impact sur l'épissage de l'ARN de variants de signification inconnue de gènes impliqués dans la prédisposition héréditaire aux cancers du sein et/ou de l'ovaire.

JURY

Président : M. le Dr Christophe DENOYELLE
Membres : Mme le Dr Pascaline BERTHET
Mme le Dr Alexandra MARTINS
Mme le Dr Sophie KRIEGER

Remerciements

Tout d'abord, je tiens à remercier Mesdames Pascaline Berthet et Alexandra Martins ainsi que Monsieur Christophe Denoyelle, pour me faire l'honneur de juger ce travail.

Je remercie M. Dominique Vaur, directeur du laboratoire de Biologie et Génétique du Cancer du Centre François Baclesse, pour ces quatre années passées dans son service de génétique. J'ai pu profiter de vos conseils, avis et remarques pour avancer pendant ces 4 années. Je remercie aussi M. Laurent Castéra, pour son regard critique et constructif sur l'avancée des travaux.

Je remercie infiniment mon encadrante de thèse Sophie Krieger. Merci pour ces 4 années, depuis le Master 2, jusqu'à aujourd'hui. Merci pour ta patience, ton énergie et ton attention afin de structurer ce projet.

Je remercie bien évidement l'ensemble du LBGC, qui m'a supporté pendant 4 ans. Merci pour votre accueil et pour avoir permis mon intégration au sein du laboratoire. Je remercie plus particulièrement Jean-Jacques, Angélina, Danielle et Antoine pour m'avoir formé aux analyses de l'ARN. Merci aussi à Antoine et Nicolas pour leur implication dans ce projet. Enfin merci au petit nouveau Raphaël pour son aide précieuse.

Merci à l'ensemble de l'équipe ARN de l'Unité Inserm 1079 : Alexandra Martins, Pascaline Gaildrat, Hélène Tubeuf, Omar Soukarieh, Thierry Frébourg, pour leur accueil, leurs échanges, et pour m'avoir formé aux tests fonctionnels minigènes.

Merci à mes parents pour leur confiance, leur amour, leur soutien tout au long de mes études. Je n'aurais pas pu réussir sans vous. Merci pour tout Maman. Je pense particulièrement à mon père, parti bien trop tôt et à qui j'aurais été tellement fier de présenter l'aboutissement de ces études passionnantes. Merci à Claire et Stan, mes petits loups, pour votre présence bienveillante. Merci à toute ma « grande » famille ainsi qu'à ma « petite » belle-famille.

Enfin, je remercie ma femme, pour son soutien inébranlable depuis le premier jour de notre rencontre. Tu es mon roc, mon amour. La vie est belle grâce à toi. L'année 2016 aura été particulièrement intense, merci de l'avoir rendue magique. J'attends avec impatience l'année 2017. Je t'aime.

Table des matières

INTRODUCTION **10**

A. PREDISPOSITION HEREDITAIRE AUX CANCERS DU SEIN ET DE L'OVAIRE **11**

I. Généralités sur les cancers du sein et de l'ovaire **11**
1. Epidémiologie des cancers du sein 11
2. Epidémiologie des cancers de l'ovaire 12

II. Prédisposition héréditaire aux cancers du sein et de l'ovaire **14**
1. Gènes de prédisposition aux cancers du sein et/ou de l'ovaire 14
   a. Le gène *BRCA1* et la protéine BRCA1 18
   b. Le gène *BRCA2* et la protéine BRCA2 21
   c. Rôles de BRCA1 et BRCA2 dans la réparation des cassures double-brin par recombinaison homologue 24
2. Problématique d'interprétation biologique et de classification des variations dans le syndrome HBOC 28

B. LE ROLE DE L'EPISSAGE DE L'ARN DANS LE SYNDROME HBOC **35**

I. L'épissage du pré-ARNm **35**
1. Les principaux signaux d'épissage 35
2. La réaction d'épissage 38
3. Le spliceosome 40
   a. Le spliceosome majeur 40
   b. Le spliceosome mineur 45
4. Régulation de l'épissage 47
   a. Eléments activateurs de l'épissage : ESE, ISE et protéines SR 48
   b. Eléments inhibiteurs de l'épissage : ESS, ISS et protéines hnRNP 49
5. Epissage alternatif de l'ARNm 50
   a. Différents types d'épissage alternatif 51
   b. Régulation de l'épissage alternatif 53

II. Epissages alternatifs dans les gènes *BRCA1* et *BRCA2* **55**
1. *BRCA1* 55
2. *BRCA2* 59

III. Anomalies de l'épissage et maladies génétiques **62**
1. Altérations des signaux d'épissage 64
   a. Altérations au niveau des sites consensus d'épissage 65
   b. Altérations au niveau des sites régulateurs de l'épissage 69
2. Altérations des facteurs régulateurs de l'épissage 70
   a. Altérations de la biogénèse des RNPsn 71
   b. Altérations des protéines composant le spliceosome 73
3. Expansion de répétitions de nucléotides – ARN « toxique » 76
4. Implication de l'épissage dans la prédisposition héréditaire aux cancers du sein et de l'ovaire 79

IV. Analyse des effets sur l'épissage des variations dans les gènes du syndrome HBOC 82
1. Prédictions bio-informatiques 82
2. Etudes expérimentales basées sur l'analyse de l'ARN de patientes atteintes du syndrome HBOC 86
3. Tests fonctionnels basés sur l'utilisation de minigènes appliqués aux gènes *BRCA1* et *BRCA2* 89
4. Interprétation de l'impact sur l'épissage des VSI 91

C. OBJECTIFS DES TRAVAUX DE THESE 93

MATERIEL ET METHODES 96

RESULTATS 107

DISCUSSION ET PERSPECTIVES 116

REFERENCES BIBLIOGRAPHIQUES 123

ANNEXES 143

Liste des Figures

	Page
Figure 1 : Incidence des cancers du sein et mortalité estimées en France entre 1980 et 2012	**11**
Figure 2 : Incidence des cancers de l'ovaire et mortalité estimées en France entre 1980 et 2012	**13**
Figure 3 : Domaines fonctionnels de la protéine BRCA1	**19**
Figure 4 : Domaines fonctionnels de la protéine BRCA2	**22**
Figure 5 : Schémas simplifiés de la réparation des cassures double-brin (DSB) par recombinaison homologue (HR) et par la jonction des extrémités non homologues (NHEJ)	**25**
Figure 6 : BRCA1 et BRCA2 ont des rôles distincts dans la réparation homologue	**27**
Figure 7 : Séquences consensus d'épissage dans les introns majeurs (A) et les introns mineurs (B) des pré-mRNA chez l'homme	**36**
Figure 8 : La réaction d'épissage du pré-mRNA	**39**
Figure 9 : Complexes dits de définition d'exon (a) et de définition d'intron (b)	**42**
Figure 10 : Les étapes d'assemblage du spliceosome majeur	**44**
Figure 11 : Etapes d'assemblage du spliceosome mineur	**46**
Figure 12 : Les différents éléments régulateurs de l'épissage	**47**
Figure 13 : Différents types d'épissages alternatifs	**52**
Figure 14 : Différents transcrits alternatifs en phase identifiés dans le gène *BRCA1*	**56**
Figure 15 : Structure de certains transcrits alternatifs de *BRCA2* contenant un site cryptique donneur/accepteur d'épissage	**60**
Figure 16 : Mécanismes d'altération de l'épissage dans les maladies génétiques	**63**
Figure 17 : Distribution des mutations ponctuelles causales répertoriées aux niveaux des sites donneur et accepteur d'épissage dans la base de données HGMD en 2006	**66**
Figure 18 : Procédure pour l'analyse des scores obtenus par les algorithmes de prédictions bio-informatiques MES et SSF dédiés aux sites d'épissage	**85**
Figure 19 : Représentation schématique des constructions minigènes pCAS2 utilisé lors du test fonctionnel.	**111**
Figure 20 : Analyse des profils d'épissage des minigènes sauvages et mutants.	**112**
Figure 21 : Analyse par RT-PCR des transcrits des gènes *RAD51C* et *CHEK2* à partir de l'ARN de patients, issu de sang périphérique collecté sur tubes PAXgene®.	**113**

Liste des Tableaux

	Page
Tableau 1 : Liste des variants étudiées pour les tests fonctionnels basés sur l'utilisation de minigènes.	**97**
Tableau 2 : Caractéristiques personnelles et familiales des patients porteurs de variants introniques impactant potentiellement l'épissage de l'ARN par les prédictions bio-informatiques (MES, SSF).	**110**

Abréviations

ADN : Acide DésoxyriboNucléique
ARNm : Acide RiboNucléique messager
ARNr : Acide RiboNucléique ribosomal
ATM : *Ataxia Telangiectasia Mutated*
BARD1 : *BRCA1 Associated RING Domain 1*
BRCA1 : *BReast CAncer 1, early onset*
BRCA2 : *BReast CAncer 2, early onset*
BRCT : *BRCA1 C-terminal domain*
BRIP1 : *BRCA1 Interacting Protein*
CDH1 : *CaDHerin 1*
CHEK2 : *CHEckpoint Kinase 2*
DSB : *Double Strand Break*
EGF : *Epidermal Growth Factor*
ENIGMA : *Evidence-based Network for the Interpretation of Germline Mutant Alleles*
ESE : *Exonic Splicing Enhancer*
ESR : *Exonic Splicing Regulator*
ESS : *Exonic Splicing Silencer*
FA : *Fanconi Anemia*
FTDP-17 : *FrontoTemporal Dementia and Parkinsonism linked to chromosome 17*
FUS : *FUsed in Sarcoma*
GATK : *Genome Analysis ToolKit*
GGC : Groupe Génétique et Cancer
HBOC : *Hereditary Breast and Ovarian Cancer*
HGMD : *Human Gene Mutation Database*
hnRNP : *heterogeneous nuclear RiboNucleoProtein*
HR : *Homologuous Recombination*
HTT : Huntingtine
INCa : Institut National du Cancer
ISE : *Intronic Splicing Enhancer*
ISR : *Intronic Splicing Regulator*
ISS : *Intronic Splicing Silencer*
kb : kilobase
LLB : Lignée LymphoBlastoïde
LS : *Lynch syndrome*
mBBP : *Branch point Binding Protein*
MMR : *MisMatch Repair*
NES : *Nuclear Exportation Sequence*
NGS : *Next Generation Sequencing*
NHEJ : *Non-Homologuous End-Joining*
NMD : *Nonsense Mediated mRNA Decay*

OB : *Oligonucleotide Binding*
PALB2 : *Partner And Localizer of BRCA2*
Protéine SR : Protéine Sérine/Arginine
PTB : *Polypyrimidine Tract Binding protein*
PTC : *Premature Termination Codon*
PTEN : *Phosphatase tensin homolog on chromosome TEN*
RAD51 : *RAD51 homolog (RecA homolog, E. coli)*
RAD51B : *RAD51 paralog B*
RAD51C : *RAD51 paralog C*
RAD51D : *RAD51 paralog D*
RBM : *RNA Binding Motif protein*
RIN : *RNA Integrity Number*
RING : *Really Interesting New Gene*
ROC : *Receiver-Operating Characteristic*
Ron : Récepteur d'origine nantais
RP : *Retinitis Pigmentosa*
RT-PCR : *Reverse Transcription – Polymerase Chain Reaction*
SF1 : *Splicing Factor 1*
SLA : Sclérose Latérale Amyotrophique
SMA : *Spinal Muscular Atrophy*
SMD : Syndrome MyéloDysplasique
SMN : *Survival of Motor Neuron*
SNP : *Single Nucleotide Polymorphism*
snRNA : *small nuclear RNA*
snRNP : *small nuclear RiboNucleoProtein*
SRE : *Splicing Regulatory Elements*
SSF : *Splice Site Finder*
STK11 : *Serine/Threonine Kinase 11*
TDP-43 : *Transactive response DNA binding Protein 43 kDa*
TP53 : *Tumor protein p53*
U2AF : *U2 Auxiliary Factor 2*
VSI : Variant de Signification Inconnue
XRCC2 : *X-ray Repair Complementing defective repair in Chinese hamster cells 2*
XRCC3 : *X-ray Repair Complementing defective repair in Chinese hamster cells 3*

Introduction

A. Prédisposition héréditaire aux cancers du sein et de l'ovaire

I. Généralités sur les cancers du sein et de l'ovaire

1. Epidémiologie des cancers du sein

Chez la femme, le cancer du sein est le cancer le plus fréquent, avec environ 54 000 nouveaux cas estimés en 2015 en France métropolitaine, soit 31% des nouveaux cas de cancers diagnostiqués annuellement chez les femmes (Binder-Foucard et al. 2014). L'âge médian au diagnostic de cancer du sein se situe entre 50 et 64 ans, et 80 % de ces cancers sont diagnostiqués après l'âge de 50 ans. On estime aujourd'hui qu'environ une femme sur 9 développera un cancer du sein au cours de sa vie. Le nombre annuel de cas de cancer du sein chez la femme en France a constamment augmenté ces 20 dernières années, passant de 30 000 cas en 1990 à 53 000 cas en 2011. Sur cette même période, l'incidence des cas de cancer du sein chez la femme en France a augmenté, passant de 75,3 cas pour 100 000 femmes en 1990 à 97,8 cas pour 100 000 femmes en 2005 (**Figure 1**). Cette incidence a tendance à diminuer depuis, avec une incidence estimée à 88 cas pour 100 000 femmes en 2012.

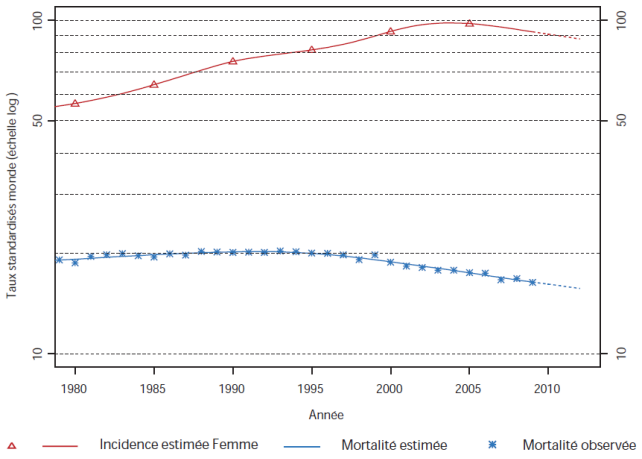

Figure 1 : Incidence des cancers du sein et mortalité estimées en France entre 1980 et 2012 (données de l'Institut National du Cancer (INCa))

Avec environ 11 900 décès estimés en 2015, le cancer du sein est au premier rang (18,2%) des décès par cancer dans la population féminine. La mortalité, est restée relativement stable jusqu'aux alentours de 1995 malgré une forte augmentation de l'incidence durant cette période, puis a diminué significativement ensuite, passant de 20,2 décès pour 100 000 femmes en 1990 à 15,7 décès pour 100 000 femmes en 2012 (**Figure 1**).

2. Epidémiologie des cancers de l'ovaire

Le cancer de l'ovaire représente en France 4 620 nouveaux cas en 2012 et 3 140 décès, ce qui en fait le cancer gynécologique de plus mauvais pronostic (Binder-Foucard et al. 2014). Il concerne principalement les femmes de plus de 50 ans (13 % des femmes sont diagnostiquées avant 50 ans). On estime aujourd'hui qu'environ une femme sur 100 développera un cancer de l'ovaire au cours de sa vie.

L'incidence des cas de cancer de l'ovaire chez la femme en France diminue depuis 20 ans, passant de 9,2 cas pour 100 000 femmes en 1990 à 7,6 cas pour 100 000 femmes en 2012 (**Figure 2**). De la même manière la mortalité estimée diminue depuis 20 ans, passant de 6 décès pour 100 000 femmes en 1990 à 3,8 décès pour 100 000 femmes en 2012 (**Figure 2**). Cette diminution du taux de mortalité serait notamment liée à la baisse récente de l'incidence de ce cancer en lien avec l'utilisation, par les générations de femmes nées après 1960, de contraceptifs oraux connus pour avoir un effet favorable à long terme sur les risques de cancers ovariens (Binder-Foucard et al. 2014).

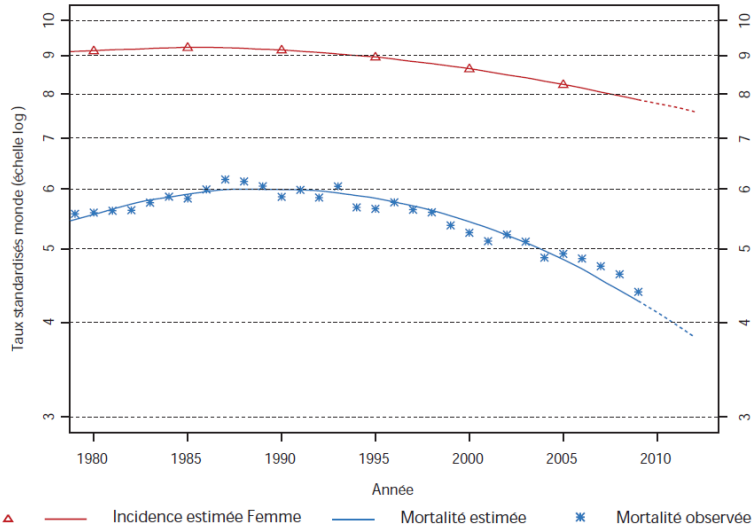

Figure 2 Incidence des cancers de l'ovaire et mortalité estimées en France entre 1980 et 2012 (données INCa)

II. Prédisposition héréditaire aux cancers du sein et de l'ovaire

1. Gènes de prédisposition aux cancers du sein et/ou de l'ovaire

Dès la fin du XIXe siècle, Paul Broca étudie l'histoire médicale d'une famille présentant 16 cas de mort par cancer sur 4 générations, en particulier 10 cas de cancer du sein (Broca 1866). Cent ans plus tard, Henry Lynch observe aussi qu'une agrégation de cas de cancers du sein et de l'ovaire (Lynch and Krush 1971; Lynch et al. 1974) semble souvent apparaitre à un jeune âge dans certaines familles (Lynch et al. 1976). Dans les années 1990, l'analyse de ce genre de familles semblait suggérer l'existence d'un (de) gène(s) de susceptibilité au cancer du sein et/ou de l'ovaire se transmettant selon un mode autosomique dominant et présentant une pénétrance élevée (Claus et al. 1991). En 1994 et 1995 ont été respectivement mis en évidence les gènes *BReast CAncer 1* (*BRCA1*) et *BReast CAncer 2* (*BRCA2*) (Miki et al. 1994; Wooster et al. 1995; Tavtigian et al. 1996).

La prédisposition héréditaire aux cancers du sein et/ou de l'ovaire (appelé par la suite syndrome HBOC pour *Hereditary Breast and/or Ovarian Cancer*) est caractérisée au sein d'une famille par la présence d'une agrégation de cas de cancers du sein diagnostiqués avant la ménopause dans une même branche parentale, ainsi qu'éventuellement un ou plusieurs cas de cancers de l'ovaire, de cancers du sein bilatéraux, de cancers du sein chez l'homme, et la présence chez une même femme de cancers du sein et de l'ovaire. Environ cinq à dix pour cent des cancers du sein et/ou de l'ovaire surviennent sur un terrain de prédisposition héréditaire (Kobayashi et al. 2013). Annuellement, en France, on estime que 2 000 nouveaux

cas de cancer du sein et 200 nouveaux cas de cancer de l'ovaire seraient liés à une altération génétique constitutionnelle (Eisinger et al. 2006).

Le syndrome HBOC, à transmission autosomique dominante, est dû dans 25 à 30% des cas à une mutation constitutionnelle dans le gène *BRCA1* ou *BRCA2* (Kast et al. 2016). La présence d'une mutation délétère dans un de ces deux gènes augmente considérablement le risque de développer un cancer du sein ou de l'ovaire. Mais la pénétrance est incomplète car le risque est inférieur à 100%. En effet, d'après l'analyse prospective de Mavaddat et collaborateurs, le risque cumulé à l'âge de 70 ans pour une femme de développer un cancer du sein est de 60% pour le gène *BRCA1* et de 55% pour le gène *BRCA2* (Mavaddat et al. 2013). Quant au risque cumulé à l'âge de 70 ans pour une femme de développer un cancer de l'ovaire, il est de 59 % pour le gène *BRCA1* et de 16.5% pour le gène *BRCA2*. Les hommes porteurs d'une mutation dans les gènes *BRCA1* et *BRCA2* présentent aussi un risque plus élevé de développer un cancer du sein. *BRCA2* est associé à la prédisposition au cancer chez l'homme d'une manière plus importante que *BRCA1*. Les hommes porteurs d'une mutation dans le gène *BRCA2* ont un risque de développer un cancer du sein de 6,8% à l'âge de 80 ans, plus élevé que celui de la population masculine générale qui est estimé à 0,12% (Tai et al. 2007). Quant aux hommes porteurs d'une mutation dans le gène *BRCA1*, ils ont un risque de 1,2% de développer un cancer du sein. De plus, les mutations au niveau des gènes *BRCA1* et *BRCA2* augmentent le risque de cancer de la prostate. En effet, les hommes porteurs d'une mutation dans *BRCA2* ont 20% de risque de développer un cancer de la prostate à l'âge de 80 ans. Ceux porteurs d'une mutation dans *BRCA1* ont un risque légèrement plus important que celui de la population générale qui est de 16% (Smith 2012).

Cependant, entre 10 et 25% des cas de HBOC peuvent être attribués à des mutations de *BRCA1* et *BRCA2* (Institut National du Cancer 2016; Kast et al. 2016). Cette hérédité manquante peut s'expliquer aujourd'hui par la présence de mutations dans d'autres gènes impliqués dans ce syndrome. Il a été montré que d'autres gènes étaient impliqués dans le déterminisme génétique des cancers du sein et/ou de l'ovaire. Par exemple, des mutations dans les gènes *TP53*, *PTEN*, *STK11* et *CDH1*, responsables respectivement du syndrome de Li-Fraumeni, du syndrome de Cowden, du syndrome de Peutz-Jeghers et du syndrome de cancer gastrique héréditaire diffus, sont associées à un risque augmenté de cancer du sein (Giardiello et al. 2000; Pharoah et al. 2001; Walsh et al. 2006; Bubien et al. 2013). Egalement, une évolution des connaissances a permis de montrer que des mutations dans le gène *PALB2* confèrent un risque similaire à *BRCA2* de développer un cancer du sein, puisque le risque cumulé de cancer du sein chez des femmes porteuses de mutations dans le gène *PALB2* est estimé aujourd'hui à 35% à 70 ans (Antoniou et al. 2014). Des mutations dans les paralogues du gène *RAD51*, comme *RAD51C*, confèrent un risque augmenté de cancer de l'ovaire (Meindl et al. 2010). Des mutations dans les gènes *ATM*, *BRIP1*, et *CHEK2* sont associées à des risques plus modérés de cancer du sein (Byrnes et al. 2008). Enfin, des mutations dans d'autres gènes peuvent être associées à un risque faible de cancer du sein ou de l'ovaire. Parmi elles, des mutations dans d'autres paralogues du gène *RAD51* pourraient aussi expliquer des cas de cancers héréditaires du sein (*RAD51B*, *XRCC2*, *XRCC3*) (Hilbers et al. 2012; Park et al. 2012; Golmard et al. 2013) ou de l'ovaire (*RAD51D*) (Loveday et al. 2012). Mais leur contribution respective et la pénétrance de ces mutations restent à caractériser.

Par la suite, nous nous focaliserons sur la description et le rôle des premiers gènes identifiés, *BRCA1* et *BRCA2*. Ces gènes sont exprimés de façon ubiquitaire chez l'humain,

avec une expression corrélée avec les niveaux de prolifération cellulaire. L'expression est de plus variable au cours du cycle cellulaire, avec des niveaux maximums à la transition G1/S (Chodosh 1998). Ces gènes codent chacun une phosphoprotéine nucléaire jouant un rôle dans le maintien de la stabilité génomique en intervenant notamment dans la réparation des cassures double-brin de l'acide désoxyribonucléique (ADN) (DSB, *double-strand break*) par recombinaison homologue (HR, *homologous recombination*). Ce sont des gènes suppresseurs de tumeurs (*caretakers*). Leur inactivation crée un état permissif dans lequel la cellule accumule des défauts génétiques amenant à une instabilité chromosomique. Bien que les fonctions précises de BRCA1 et BRCA2 ne soient pas entièrement comprises, il est clair que ce sont des protéines multifonctionnelles. Les protéines BRCA1 et BRCA2 jouent également des rôles importants dans la régulation de la transcription, le contrôle du cycle cellulaire et le remodelage de la chromatine.

a. Le gène *BRCA1* et la protéine BRCA1

Le gène *BRCA1* est situé sur le chromosome 17 (en 17q21). Il est composé de 23 exons, dont 22 codants, distribués sur plus de 80 kilobases (kb). La transcription de ce gène produit un transcrit principal de 7.2 kb qui est ensuite traduit en une protéine de 1 863 acides aminés. Il est caractérisé par un exon 11 représentant environ 50% de la séquence codante. Des transcrits plus courts, dits alternatifs, ont été décrits (Colombo et al. 2014). Les rôles précis de ces transcrits alternatifs restent à préciser.

La protéine comprend un domaine catalytique RING (*Really interesting new gene*) en N-terminal (**Figure 3**). Le domaine RING est composé d'une répétition d'un motif de 8 acides aminés conservés (Cys3-His-Cys4) et est impliqué dans l'ubiquitination de la protéine et les interactions protéine-protéine. Ce domaine interagit avec la protéine BARD1 (*BRCA1 associated RING domain 1*) pour former un complexe hétérodimérique à activité ubiquitine ligase E3. Cette interaction BRCA1/BARD1 est d'ailleurs essentielle pour la stabilité des deux protéines (Wu et al. 2008). Les enzymes à activité ubiquitine ligase E3 sont responsable de l'ubiquitination d'un grand nombre de protéines, y compris les histones et la protéine BRCA1 elle-même (Wu et al. 2008). Cette ubiquitination permet de marquer les protéines à dégrader par le complexe protéasome. En plus de son activité ubiquitine ligase E3, l'interaction BRCA1/BARD1 est importante pour le rôle de BRCA1 dans la suppression de tumeurs et la réparation de l'ADN (Savage and Harkin 2015).

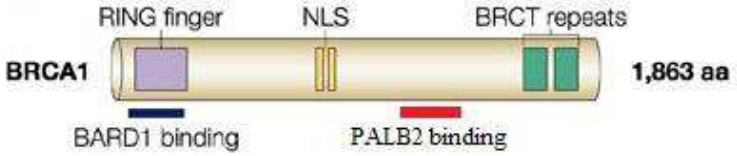

Figure 3 : Domaines fonctionnels de la protéine BRCA1 (d'après West 2003). Certaines des protéines interagissant avec BRCA1 sont indiquées en-dessous du site d'interaction. RING : *Really interesting new gene* ; NLS : *nuclear localization signal* ; BRCT : *BRCA1 C-terminal domain*.

La protéine comprend aussi 2 copies en tandem du domaine BRCT (*BRCA1 C-terminal domain*) (**Figure 3**). Le domaine BRCT est impliqué dans des interactions protéiques et des études suggèrent que les domaines BRCT en tandem constituent un motif de reconnaissance et de liaison de peptides phosphorylés (Manke et al. 2003; Yu et al. 2003). Ces domaines jouent aussi un rôle dans l'activation de la transcription par BRCA1 (transactivation) (Chapman and Verma 1996; Monteiro et al. 1996).

Bien qu'elle soit une protéine nucléaire, possédant deux séquences de localisation nucléaire (NLS) (**Figure 3**) dans la portion centrale de la protéine (acides aminés 503-509 et 606-615) (Thakur et al. 1997), BRCA1 a été aussi identifiée au niveau du cytoplasme et de la mitochondrie (Coene et al. 2005). En dehors du noyau, BRCA1 régule la division cellulaire, les réarrangements du cytosquelette, l'apoptose et la réparation du génome mitochondrial (Hsu and White 1998; Lingle et al. 1998; Deng and Brodie 2000). Cette migration entre les compartiments cellulaires est modulée par une séquence d'exportation nucléaire (NES) retrouvée à l'extrémité du domaine RING (acides aminés 81-99) (Rodríguez and Henderson 2000).

La région centrale de la protéine BRCA1 est codée en grande partie par l'exon 11 composé de 3 426 nucléotides (nt). La région codée par cet exon ne possède pas de domaine fonctionnel connu. La protéine interagit avec de nombreuses protéines et contient plusieurs sites de phosphorylation en réponse aux dommages de l'ADN, qui interviennent dans la formation des différents complexes protéiques de BRCA1 (Savage and Harkin 2015). Elle est le substrat de plusieurs kinases cellulaires, dont ATM (*Ataxia Telangiectasia Mutated*) et CHEK2 (*Checkpoint Kinase 2*) dont les gènes comportent des altérations qui ont aussi été associées à une susceptibilité accrue au cancer du sein. La kinase la mieux caractérisée de BRCA1 est ATM, qui phosphoryle la protéine sur plusieurs résidus sérine (Ser 1189, 1298, 1330, 1387, 1423, 1457, 1466, 1524 et 1542) (Cortez et al. 1999), bien que d'autres kinases phosphorylent aussi la protéine (Ouchi 2006). La protéine BRCA1 contient un motif superhélice qui lui permet d'interagir avec BRCA2 par l'intermédiaire de PALB2 (*Partner and Localizer of BRCA2*) (Zhang et al. 2009) (**Figure 3**).

BRCA1 intervient également dans plusieurs mécanismes cellulaires essentiels pour la stabilité du génome comme la régulation du cycle cellulaire, le remodelage de la chromatine et l'apoptose (Savage and Harkin 2015). L'invalidation à l'état homozygote du gène *BRCA1* est létal chez la souris au stade embryonnaire (Gowen et al. 1996), suggérant que *BRCA1* est impliqué dans la prolifération et la différenciation des cellules souches (Hakem et al. 1996).

b. Le gène *BRCA2* et la protéine BRCA2

Le gène *BRCA2* est situé sur le chromosome 13 (en 13q12.3). Il est composé de 27 exons, dont 26 codants, distribués sur environ 84 kilobases (kb). La transcription de ce gène produit un transcrit principal de 11.4 kb qui est ensuite traduit en une très grosse protéine de 3 418 acides aminés. Il est lui aussi caractérisé par un exon 11 représentant environ 50% de la séquence codante. Des transcrits alternatifs ont été décrits (Fackenthal et al. 2016). Les rôles précis de ces transcrits alternatifs restent à préciser.

La région centrale et l'extrémité C-terminale de la protéine BRCA2 sont les mieux caractérisées. BRCA2 présente très peu d'homologie avec d'autres protéines, et aucune avec BRCAl. L'analyse de la séquence protéique a permis de mettre en évidence la présence d'un motif (appelé BRC) répété 8 fois dans la région centrale de BRCA2 (Bork et al. 1996) (**Figure 4**). Une avancée majeure dans la compréhension des fonctions cellulaires de BRCA2 est venue de la réalisation que l'interaction entre BRCA2 et RAD51 est médiée par ces répétitions (Wong et al. 1997; Chen et al. 1998). Les protéines BRCA2 et RAD51 interviennent ensemble dans la recombinaison homologue des cassures double-brin de l'ADN.

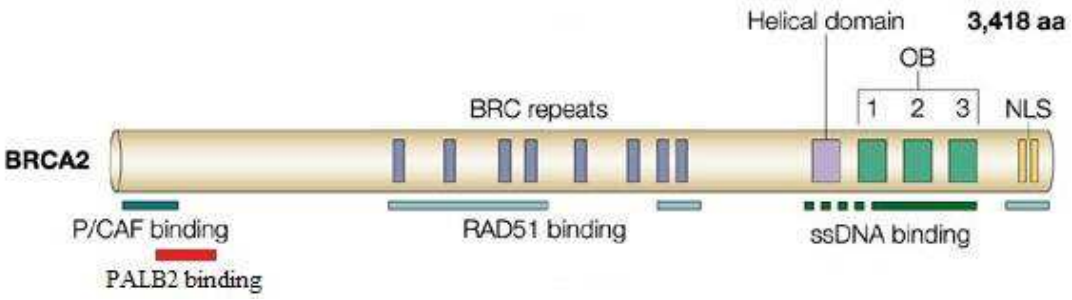

Figure 4 : Domaines fonctionnels de la protéine BRCA2 (d'après West 2003). Certaines des protéines interagissant avec BRCA1 sont indiquées en-dessous du site d'interaction. NLS : *nuclear localization signal* ; BRCT : *BRCA1 C-terminal domain* ; OB : *oligonucleotide binding*.

La portion C-terminale de BRCA2 comprend une région d'environ 800 acides aminés, appelée DBD (*DNA binding domain*). Le premier motif de cette région est appelé le domaine hélical et comprend 190 acides aminés qui forment principalement des hélices α (**Figure 4**). Puis viennent trois domaines OB (*oligonucleotide binding*) d'environ 110 acides aminés (OB1, OB2 et OB3) qui permettent la liaison à l'ADN simple-brin (**Figure 4**). La région C-terminale contient aussi deux NLS (Yano et al. 2000) ainsi qu'un site de phosphorylation par des kinases dépendantes des cyclines (CDK; *cyclin-dependent kinases*) semblant réguler l'association de RAD51 à un deuxième motif de liaison à cet endroit (Esashi et al. 2005) (**Figure 4**).

La protéine BRCA2 possède un domaine de liaison à la protéine PALB2 au niveau de son extrémité N-terminale (Xia et al. 2006) (**Figure 4**). L'interaction avec PALB2 est essentielle pour l'activité de recombinaison homologue. Dans la même région il a été montré que le domaine codé par l'exon 3 du gène *BRCA2* était impliqué dans l'activation de la transcription (Milner et al. 1997). Si ses cibles ne sont pas encore toutes bien caractérisées, cette fonction semble être régulée par plusieurs interactions. Par exemple, l'interaction de

BRCA2 avec le co-activateur de transcription P/CAF au niveau de cette région intervient dans la fonction acétyltransférase de BRCA2 (Fuks et al. 1998) (**Figure 4**). De plus, BRCA2 interagit avec BRCA1 et P/CAF pour activer la transcription du récepteur aux androgènes et GRIP1 (un co-activateur des récepteurs aux androgènes) afin d'initier l'effet antiprolifératif de ces récepteurs (Shin and Verma 2003). Un autre aspect de l'implication de BRCA2 dans la transcription a été mis en évidence par la découverte d'un nouveau partenaire d'interaction, EMSY, qui semble agir comme répresseur de la transcription (Hughes-Davies et al. 2003). De façon intéressante il a été démontré qu'EMSY est amplifié dans un nombre significatif de tumeurs du sein sporadiques et est associé à un mauvais pronostic (Hughes-Davies et al. 2003).

L'invalidation à l'état homozygote du gène *BRCA2* est létal au stade embryonnaire chez la souris (Sharan et al. 1997), ce qui met en évidence l'importance de la protéine BRCA2 dans le maintien de la stabilité du génome. BRCA2 intervient dans plusieurs autres mécanismes cellulaires comme la maintenance des télomères, le métabolisme des boucles en R ADN-ARN et la mitose (Martinez et al. 2015).

c. Rôles de BRCA1 et BRCA2 dans la réparation des cassures double-brin par recombinaison homologue

Les cellules déficientes en BRCAl ou BRCA2 montrent une inefficacité à effectuer la réparation des cassures de l'ADN et en particulier les lésions impliquant des cassures double-brin (Zhang and Powell 2005; Thorslund and West 2007). Elles peuvent être dues à des erreurs de réplication, à l'exposition à des radiations ionisantes ou à d'autres composés génotoxiques. En fonction du type cellulaire ou du moment du cycle, la cellule dispose de voies de réparation par recombinaison homologue (HR) ou par la jonction des extrémités non-homologues (NHEJ, *non-homologous end-joining*) pour réparer les cassures double-brin (**Figure 5**). La voie NHEJ ne nécessite pas d'homologie. Cette voie peut résulter en une perte ou une addition de nucléotides aux sites de réparation et est habituellement considérée comme introduisant des erreurs, pouvant entraîner une instabilité du génome. La réparation homologue, qui nécessite une chromatide sœur intacte comme modèle, ne peut par conséquent survenir que dans les phases cellulaires S et G2 (Roy et al. 2012). Elle est considérée comme une voie de réparation fiable et la plus efficace.

Les protéines BRCAl et BRCA2 ont des rôles distincts dans la réparation homologue. La protéine BRCAl joue majoritairement un rôle de médiateur entre les protéines de détection des cassures et les protéines impliquées dans la réparation. Le principal rôle de BRCA2 consiste à réguler l'action de la recombinase RAD51 impliquée dans la réparation homologue (Venkitaraman 2009).

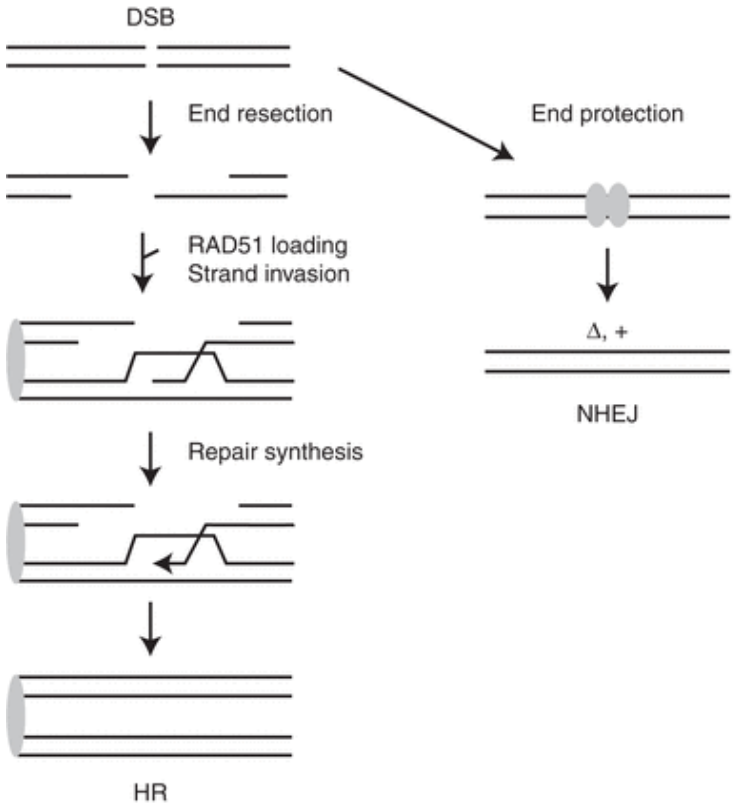

Figure 5 : Schémas simplifiés de la réparation des cassures double-brin (DSB) par recombinaison homologue (HR) et par la jonction des extrémités non homologues (NHEJ) (d'après Prakash et al. 2015). Par la recombinaison homologue, les 2 extrémités sont coupées par des protéines de résection. La protéine RAD51 se fixe et envahit la chromatide sœur afin de l'utiliser comme modèle pour la réparation. L'extrémité 3' amorce la synthèse de l'ADN pour restaurer la séquence d'origine. Enfin, le brin néo-synthétisé sert de matrice pour la réparation du 2e brin. Par la NHEJ, les extrémités d'ADN sont protégées d'une résection par des protéines NHEJ, les étapes suivantes pouvant entraîner une mauvaise réparation avec des insertions (+) ou des délétions (Δ) de nucléotides.

Bien que ces fonctions soient moins bien comprises que celles de BRCA2, BRCAl est aussi impliquée dans les voies de réparation des cassures double-brin. Le recrutement des protéines de détection ATM (*ataxia telangiectasia mutated*), ATR (*ataxia telangiectasia and Rad3-related*) et de la nibrine (NBS1) est l'un des premiers événements ayant lieu au niveau du site de cassure, ce qui permet l'activation de la kinase ATM (Lee and Paull 2005). ATM permet alors l'accumulation de plusieurs protéines au niveau du point de cassure, incluant BRCAl et le complexe MRE11A/RAD50/NBS1 (MRN). À son tour, BRCAl est phosphorylée

et requise pour la phosphorylation par ATM et ATR de protéines de signalisation et de réparation (Foray et al. 2003). BRCA1 interagit ensuite via son domaine BRCT avec CtIP (complexe BRCA1-C) préalablement phosphorylé (Yu et al. 1998) (**Figure 6**). Cette interaction favorise la coopération avec le complexe MRN pour catalyser la résection de l'ADN.

La participation de BRCAl dans plusieurs étapes de la réponse aux cassures de l'ADN est attestée par ses interactions protéiques et sa présence dans plusieurs complexes distincts. BRCA1 est un antagoniste de 53BP1 (*p53 binding protein 1*), impliqué dans le blocage de la résection (Bunting et al. 2010) (**Figure 6**). En l'absence de BRCA1, 53BP1 s'accumule au niveau du point de cassure et bloque la résection, donc la recombinaison homologue, ce qui génère une aberration chromosomique et la mort cellulaire.

BRCAl est impliquée dans l'ubiquitination de protéines par l'intermédiaire de son association avec BARD1, qui augmente en retour son activité ubiquitine ligase (Xia et al. 2003). L'hétérodimère BRCAl-BARD1 est présent dans plusieurs complexes protéiques et contrôle notamment la recombinaison homologue en interagissant avec BRCA2, RAD51 et PALB2 (Greenberg 2008) (**Figure 6**). La perturbation de BRCA1 inhibe la fixation de ces 3 protéines (Sy et al. 2009).

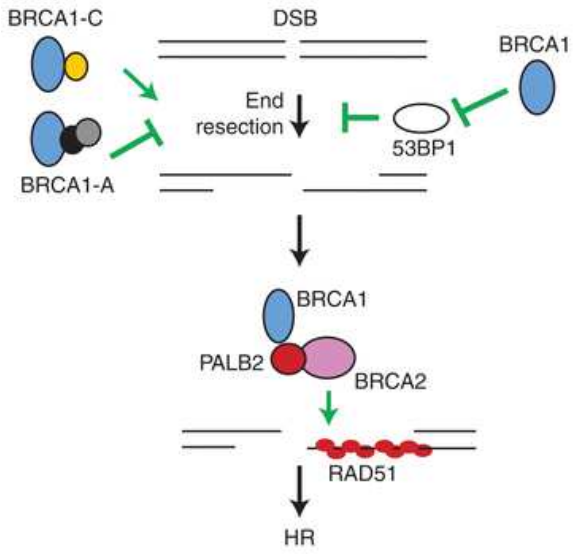

Figure 6 : BRCA1 et BRCA2 ont des rôles distincts dans la réparation homologue (Prakash et al. 2015)**.** BRCA1 agit précocement pour promouvoir la coupure des extrémités de l'ADN. Il recrute ensuite PALB2 et par conséquent BRCA2. Parallèlement, BRCA1 inhibe l'inhibiteur de résection 53BP1. Il régule la résection par le recrutement de CtIP (en jaune) via le complexe BRCA1-C. BRCA2 favorise le chargement de la recombinase RAD51 sur le produit de résection, essentiel pour la recombinaison homologue.

Quant à BRCA2, la protéine interagit avec la recombinase RAD51 (Chen et al. 1998). Elle est recrutée par BRCA1 par l'intermédiaire de la protéine PALB2 (**Figure 6**). La perturbation de PALB2 diminue la fixation de BRCA2 et RAD51 (Xia et al. 2006). À la suite de dommages à l'ADN, BRCA2 et RAD51 se colocalisent dans des foci nucléaires, ce qui suggère que BRCA2 est nécessaire à la localisation adéquate de RAD51 et à ses fonctions (Yuan et al. 1999). BRCA2 interagit avec RAD51 par ses motifs BRCs, permettant à BRCA2 de séquestrer RAD51 lorsque la réparation n'est pas nécessaire. Une fois la réparation homologue enclenchée, BRCA2 est responsable de la relocalisation de RAD51 au niveau de la cassure double-brin, ce qui permet à RAD51 d'interagir avec l'ADN. Suite à cette interaction, RAD51 permet la recherche de la séquence homologue qui servira de modèle pour la réparation.

2. Problématique d'interprétation biologique et de classification des variations dans le syndrome HBOC

Depuis l'invalidation des brevets des gènes *BRCA1* et *BRCA2*, qui donnaient à la société commerciale Myriad Genetics une situation de quasi-monopole pour les tests moléculaires sur ces gènes en Amérique du Nord, combiné à l'essor du séquençage à haut débit (NGS, *next generation sequencing*), le nombre de tests génétiques pratiqués a explosé. Le séquençage à haut débit d'un panel de gènes impliqués dans le syndrome HBOC a permis d'augmenter légèrement le taux de détection de mutations délétères chez les patients, estimé à 18% (Castéra et al. 2014), utilisables pour le conseil génétique et permettant une prise en charge adaptée des patients et des membres de sa famille. Pour les gènes *BRCA1* et *BRCA2*, très largement étudiés, les variations détectées comprennent des réarrangements génomiques de grande taille de type délétion/duplication, des insertions/délétions de quelques nucléotides (indels) pouvant conduire à un décalage du cadre de lecture (*frameshift*), des variations ponctuelles de type non-sens, faux-sens ou encore introniques (Kobayashi et al. 2013).

En parallèle, le NGS a augmenté fortement le nombre de variants détectés au sein de ces gènes. En conséquence, le défi n'est plus la détection de variations chez les patients mais leur interprétation biologique et la détermination de leur classe de pathogénicité pour le diagnostic moléculaire (Frebourg 2014; Mort et al. 2014; Wu et al. 2014). Le plus souvent, la signification biologique des mutations tronquantes, telles que des mutations non-sens et des indels, semble évidente, par apparition d'un codon STOP prématuré. Pour les autres variations (faux-sens, indels en phase, synonymes, introniques), l'interprétation de la signification biologique est beaucoup plus difficile. Elles sont alors considérées comme des

variants de signification inconnue (VSI). Dans certains cas, l'interprétation de ces VSI fait appel à l'analyse de la ségrégation du variant dans les familles incluant plusieurs cas de cancer (Frebourg 2014). Néanmoins, dans de nombreux cas, l'interprétation de ces VSI reste difficile, les rendant inutilisables pour la gestion clinique des patients et de leurs familles, tant qu'une preuve formelle de pathogénicité n'est pas apportée. Par le nombre de tests moléculaires réalisés, l'impact médical pour les patients et leurs familles de l'identification d'une mutation délétère, le pourcentage élevé de VSI détectés, les gènes *BRCA1* et *BRCA2* sont devenus le paradigme du défi de l'interprétation des VSI en génétique médicale (Frebourg 2014).

Un outil important pour évaluer la signification clinique des VSI est le modèle de vraisemblance multifactoriel (Goldgar et al. 2004; Chenevix-Trench et al. 2006; Easton et al. 2007; Tavtigian et al. 2008; Lindor et al. 2012), qui intègre des données provenant de sources multiples, y compris la bio-informatique, la clinique, la pathologie tumorale et les données moléculaires afin d'évaluer si un variant présente des caractéristiques associées à des mutations pathogènes connues. Cet outil statistique a été utilisé pour classer les variants rares dans les gènes *BRCA1* et *BRCA2*. Ce modèle a servi de base pour classer les variants en cinq classes, chacune liée à des recommandations cliniques (Plon et al. 2008). La classe 1 regroupe les variants non pathogènes ou sans signification clinique ; la classe 2 regroupe les variants probablement non pathogènes ou de faible signification clinique ; la classe 3 regroupe les variants de signification incertaine (VSI) ; la classe 4 regroupe les variants probablement pathogènes et la classe 5 regroupe les variants définitivement pathogènes (mutations délétères).

En France, l'ensemble des laboratoires labélisés par l'INCa du consortium GGC (Groupe Génétique et Cancer) travaille sur l'annotation des variations dans les gènes *BRCA1* et *BRCA2*. La base de données BRCA Share™, anciennement UMD BRCA1/BRCA2 pour le territoire français (Universal Mutation Database, http://www.umd.be/), a été développée dans le but de collecter les variations détectées dans les gènes *BRCA1* et *BRCA2* par les laboratoires français (Caputo et al. 2012; Béroud et al. 2016). Cette base de données a l'avantage de fournir des informations très utiles sur la co-occurrence éventuelle d'un variant avec d'autres variations sur le même gène ou d'autres gènes. Elle regroupe aujourd'hui plus de 6 200 variations dans les gènes *BRCA1* et *BRCA2* (Béroud et al. 2016). Dans cette base, les variations constitutionnelles identifiées et répertoriées sont majoritairement des variations ponctuelles situées au niveau de la région codante puisqu'elles représentent 49,8% et 59,1% de l'ensemble des variations des gènes *BRCA1* et *BRCA2*, respectivement. Parmi toutes les variations détectées, le taux de VSI est estimé à 56% (1423 VSI / 2543 variants uniques déposés) et 64,9% (2408 VSI / 3711 variants uniques déposés) des variations détectées respectivement dans les gènes *BRCA1* et *BRCA2* (Béroud et al. 2016).

Les bases de données de variants jouent un rôle critique dans le processus de classification, en fournissant des informations générales sur l'état des connaissances scientifiques, des évaluations cliniques d'un variant par d'autres laboratoires, des preuves de causalité et des données sur les autres patients porteurs du même variant. Dans le contexte des tests moléculaires des gènes *BRCA1* et *BRCA2*, les bases de données disponibles aujourd'hui sont nombreuses et comprennent notamment BIC (Breast cancer Information Core database) (Shen and Vadgama 1999), LOVD (Leiden Open Variation Database) (Fokkema et al. 2011), ClinVar (Landrum et al. 2014) BRCA Exchange (Global Alliance for Genomics and Health 2016), ARUP (www.arup.utah.edu/) et BRCA Share™ (Caputo et al. 2012; Béroud et al.

2016). Il est à noter que Myriad Genetics, a développé sa propre base de données dont la qualité et le contenu ne peuvent pas être évalués car la société en est propriétaire et les données sont non accessible à la communauté. Ces bases de données ont été développées avec des objectifs différents et ont des processus de curation, de qualité des données et de contenu différents, comme souligné par Vail et collaborateurs (Vail et al. 2015). Ces auteurs ont comparé les bases de données BIC, ClinVar, LOVD et UMD et ont conclu que la multiplication de ces bases de données ne permet pas l'amélioration de la classification des variants et peut même s'avérer dangereuse, en raison des nombreuses incohérences de classification des variants entre elles et parfois même au sein de la même base de données, comme l'illustre ClinVar. Ces incohérences mettent en péril la prise en charge des patients et de leur famille dans ses aspects diagnostiques, pronostiques et théranostiques.

Aujourd'hui un effort international vise à améliorer les critères de classification des variations et donc permettre finalement la bonne reclassification de ces variations. Dernièrement, le consortium international ENIGMA (*evidence-based network for the interpretation of germline mutant alleles*) (Spurdle et al. 2012), constitué dans le but de faciliter la détermination de la signification clinique des variations identifiées dans les gènes *BRCA1* et *BRCA2*, a mis en place, en 2015, des bonnes pratiques standardisées afin d'améliorer la classification de ces variations dans ces gènes (https://enigmaconsortium.org/wp-content/uploads/2016/06/ENIGMA_Rules_2015-03-26.pdf). Ces bonnes pratiques sont grandement inspirées des règles de classification utilisées par le consortium GGC. Ce système décrit différents critères pour 5 classes de variations dans les gènes *BRCA1* et *BRCA2* :

- Le système des 5 classes décrit par Plon et utilisant le modèle de vraisemblance multifactoriel cité précédemment (Lindor et al. 2012);

- Le système de 5 classes pour l'interprétation des variants impactant potentiellement l'épissage et des altérations de l'épissage développés par les collaborateurs ENIGMA calqué sur le système de classification de Plon (Walker et al. 2013) ;
- Des éléments génériques du schéma à 5 classes quantitatif/qualitatif utilisé pour la classification des variants des gènes MMR (*mismatch repair*) et développé par InSiGHT (Thompson et al. 2014). Ces critères sont aussi basés sur la nature de la variation, sa fréquence, des données fonctionnelles, des données familiales et une analyse quantitative multifactorielle basée sur des données tumorales et des données de ségrégation ;
- Des éléments génériques des lignes directrices de l'ACMG (*American College of Medical Genetics and Genomics*) pour l'interprétation des variations de séquence (Richards et al. 2008) ;
- Des critères de classification développés par les laboratoires participant à ENIGMA, y compris les réseaux nationaux établis, dont ceux utilisés par le groupe GGC français ;
- La classification des variations de séquence selon les pratiques cliniques standards.

L'étude de la conséquence biologique d'une variation consiste à rechercher une altération de la structure ou de la fonction de la protéine, à l'aide de prédictions bio-informatiques et d'analyses expérimentales. Ces dernières sont basées sur des tests fonctionnels de l'activité de la protéine *BRCA1* (Millot et al. 2012) ou *BRCA2* (Guidugli et al. 2014). Parmi les tests communs aux 2 gènes, il existe :

- les tests fonctionnels basés sur l'utilisation de cellules souches embryonnaires de souris (Kuznetsov et al. 2008; Chang et al. 2009; Bouwman et al. 2013). Ces cellules contiennent un allèle défectueux du gène d'intérêt et l'autre allèle fonctionnel mais pouvant être inactivé sous condition. Le test repose sur l'observation qu'une protéine

fonctionnelle est essentielle pour la survie de ces cellules. La transfection d'un plasmide contenant l'allèle portant la variation à tester permet d'observer ou non la survie cellulaire ;

- les tests basés sur l'activité de réparation directe par homologie (Ransburgh et al. 2010; Guidugli et al. 2013). Il repose sur la présence de 2 allèles inactifs encodant le gène de la protéine fluorescente verte (*GFP*) dans le génome de cellules déficientes en *BRCA1* ou *BRCA2*. Un plasmide contenant l'allèle portant la variation à tester est transfecté dans la cellule, puis une cassure double-brin est induite au niveau de *GFP* par l'utilisation d'une endonucléase. Si la recombinaison homologue est efficace, la cellule exprimera alors *GFP* ;
- les tests basés sur l'amplification du centrosome (Farrugia et al. 2008; Kais et al. 2012). En effet, l'inactivation des gènes *BRCA1* ou *BRCA2* provoque une amplification du nombre de centrosomes, visualisable par immunofluorescence.

Des approches ont été expérimentées afin d'établir des signatures de mutations des gènes *BRCA1* ou *BRCA2* dans les tumeurs, comme l'HRD score (*homologous recombination deficiency*). L'HRD score est développé à partir de 3 données, déterminées sur des prélèvements somatiques :

- Telomeric allelic imbalance (TAI) (Birkbak et al. 2012), correspondant à une différence d'expressivité entre 2 allèles ;
- Loss of heterozigosity (LOH) (Abkevich et al. 2012), correspondant à des pertes chromosomiques supérieures à 15 Mb, reflet d'un déficit de la recombinaison homologue ;
- Large-scale state transition (LST) (Popova et al. 2012), correspondant à des cassures chromosomiques entre 2 régions adjacentes et proches.

En fonction de l'étude de ces 3 variables, un score élevé conclut à un déficit de la recombinaison homologue. L'élévation de ce score est corrélée à la fois à la présence de mutations délétères dans les gènes *BRCA1*, *BRCA2* et l'ensemble des gènes impliqués dans la recombinaison homologue au niveau somatique et/ou constitutionnel (Telli et al. 2016). L'HRD Score pourrait donc être une aide pour classer des VSI constitutionnels dans les gènes *BRCA1*, *BRCA2* ainsi que les autres gènes de la recombinaison homologue.

Parallèlement à l'ensemble de ces critères décrits précédemment, il est important de noter que l'ensemble des VSI, mais aussi les mutations tronquantes à priori délétères par impact au niveau protéique, peut provoquer une altération de l'épissage du pré-ARNm (acide ribonucléique messager) de ces gènes, mécanisme cellulaire survenant en amont de la traduction de la protéine. Ceci complexifie encore un peu plus la classification de l'ensemble de ces variations. Des analyses de l'ARN devraient donc potentiellement être effectuées pour l'ensemble des variations détectées, même celle à priori sans impact sur la protéine (Thompson et al. 2014). Aujourd'hui il est établi que les variations qui affectent l'épissage de l'ARN ont un impact clinique et sont à l'origine de maladies génétiques (Wang and Cooper 2007). Bien que la fréquence des mutations d'épissage varie considérablement entre les gènes, il est probable qu'environ 15% des mutations pathogènes provoquent la maladie par le défaut qu'elles introduisent dans les mécanismes d'épissage (Baralle et al. 2009). De nombreux VSI pourraient donc se révéler entraîner des anomalies d'épissage, en particulier dans les cancers héréditaires du sein et/ou de l'ovaire. L'étude de l'impact sur l'épissage des VSI occupe une place importante dans l'effort de classification du GGC. La classification particulière des VSI impactant l'épissage est décrite plus loin.

B. Le rôle de l'épissage de l'ARN dans le syndrome HBOC

I. L'épissage du pré-ARNm

Dès 1977, Roberts et Sharp remarquent que la séquence d'un ARNm provenant d'un adénovirus correspondait à 4 régions distinctes de l'ADN adénoviral, ce qui leur a valu l'attribution du prix Nobel en 1993 (Carr 1993). Dans la foulée, Gilbert suggérait d'appeler « exons » les séquences exprimées et séparées par des « introns » correspondant aux régions intragéniques (Gilbert 1978).

L'épissage de l'ARNm correspond à l'élimination des introns des ARN précurseurs (pré-ARNm) et à l'assemblage des exons entre eux. L'épissage se déroule dans le noyau des cellules eucaryotes, produisant un ARNm mature qui sera exporté dans le cytoplasme pour être traduit en protéine. Dans les cellules humaines, environ 80% des réactions d'épissage se déroulent en même temps que la transcription (Girard et al. 2012). Avec la synthèse de la coiffe en 5' et la polyadénylation de l'ARNm en 3', permettant de protéger l'ARNm de la dégradation dans le cytoplasme, l'épissage constitue une des étapes clés de la maturation de l'ARNm.

1. Les principaux signaux d'épissage

Chez la majorité des eucaryotes, il existe 2 types d'introns : les introns majeurs (U2), représentant 99% des introns, et les introns mineurs (U12). A chaque extrémité de ces introns sont présentes de courtes séquences consensus, correspondant aux sites 5' et 3' d'épissage, et séparant les introns des exons adjacents (**Figure 7**). Ces séquences constituent les signaux les

plus importants avec le site de branchement (BPS, *branch point site*), situé en amont du site 3' d'épissage. Ces signaux sont reconnus par la machinerie d'épissage que constitue le spliceosome.

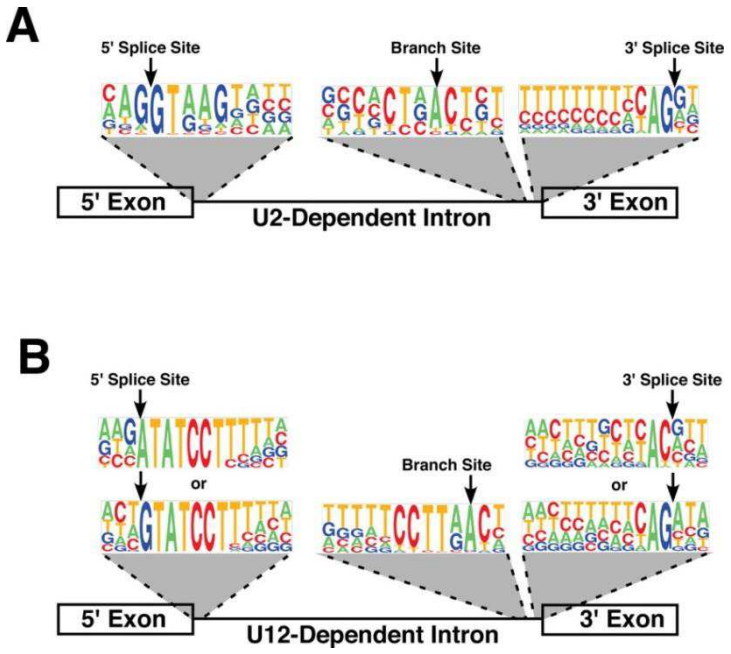

Figure 7 : Séquences consensus d'épissage dans les introns majeurs (A) et les introns mineurs (B) des pré-mRNA chez l'homme (d'après Padgett 2012). La taille des lettres reflète la fréquence d'occurrence du nucléotide correspondant à chaque position.

Dans les introns majeurs, le site 5' d'épissage se caractérise par la séquence de référence CAG|**GT**RAGT et s'étend de la position -3 à la position +8 (la numérotation négative correspond à des positions exoniques, la numérotation positive à des positions introniques ; les 2 premiers nucléotides introniques les plus conservés sont en gras, R correspond à une purine (Adénine ou Guanine)) (Cartegni et al. 2002) (**Figure 7A**). Le site 3' d'épissage se caractérise par la séquence de référence Y**AG**|G et s'étend de la position -12 à la position +2 (la numérotation négative correspond à des positions introniques, la

numérotation positive à des positions exoniques ; les 2 derniers nucléotides introniques les plus conservés sont en gras, Y correspond à une pyrimidine (Thymine (Uracile dans l'ARNm) ou Cytosine)) (Cartegni et al. 2002) (**Figure 7A**). Ces dinucléotides introniques terminaux GU et AG des sites 5' et 3' d'épissage caractérisent 99% des introns (Fredericks et al. 2015), d'où leur appellation d'introns majeurs. Le site de branchement se situe entre 18 et 40 nucléotides en amont du site 3' d'épissage, caractérisé par la séquence de référence YNYUR**A**C (le nucléotide intronique hautement conservé est en gras, N correspond à l'un quelconque des nucléotides) (Cartegni et al. 2002) (**Figure 7A**). En plus de ces sites, les introns majeurs possèdent une région riche en pyrimidines $(Py)_n$ (PPT, *polypyrimidine tract*) située entre 10 et 12 nucléotides en amont du site 3' d'épissage (Cartegni et al. 2002).

Dans les introns mineurs, les sites 5' et 3' d'épissage sont caractérisés par une combinaison différentes des dinucléotides conservés (AT ou GT en 5' ; AC ou AG en 3') (**Figure 7B**) (Dietrich et al. 1997). De plus, le site 5' d'épissage est très conservé sur les 8 à 10 premiers nucléotides de l'intron (**Figure 7B**). De même, le site de branchement, situé entre 11 et 13 nucléotides en amont du site 3' d'épissage, est très conservé. On remarque l'absence de région riche en pyrimidines à proximité du site 3' d'épissage. La grande conservation du site de branchement ainsi que sa position sont des facteurs importants pour leur identification par la machinerie d'épissage (Burge et al. 1998).

Chez l'homme, la taille moyenne des introns est comprise entre 3 000 et 4 000 nucléotides en moyenne, mais il existe également, pour les introns majeurs, de petits introns d'environ 90 nucléotides et beaucoup d'introns de très grande taille (plusieurs kb) (Patel and Steitz 2003). Enfin, les introns mineurs se retrouvent avec les introns majeurs dans les mêmes gènes, bien qu'ils soient épissés différemment. Ces introns se trouvent notamment dans des

gènes impliqués dans la réplication et la réparation de l'ADN, la transcription des ARNm et la traduction des protéines (Burge et al. 1998).

2. La réaction d'épissage

L'épissage proprement dit s'effectue en deux temps, par deux réactions de trans-estérification. On a tout d'abord une attaque nucléophile du 2'-OH du ribose de l'adénosine du site de branchement sur le phosphate de la jonction exon-intron en 5' (site 5' d'épissage) (**Figure 8**). Après cette coupure, le 3'-OH libéré au niveau de l'exon en amont attaque le phosphate de la jonction intron-exon en aval. Les produits de cette réaction sont d'une part les deux exons ligaturés correctement et d'autre part, l'intron cyclisé au niveau de l'adénosine du site de branchement. En raison de sa forme particulière, cette forme de l'intron est appelée lasso (*lariat*) (**Figure 8**). Le lasso est finalement ouvert par une enzyme de débranchement afin de pouvoir être recyclé. La bonne réalisation de ces réactions nécessite une grande précision dans l'assemblage des composants de la machinerie d'épissage, appelée spliceosome.

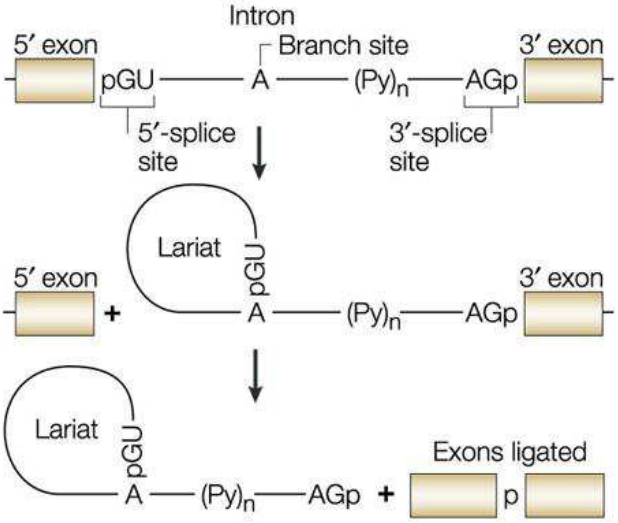

Figure 8 : La réaction d'épissage du pré-ARNm (d'après Pagani and Baralle 2004). Il existe plusieurs motifs conservés dans les séquences nucléotidiques proches des jonctions intron-exon qui agissent comme des signaux d'épissage essentiels : les dinucléotides GU et AG au niveau des jonctions exon-intron et intron-exon, respectivement (sites 5' et 3' d'épissage), une région riche en pyrimidines (Py)n et un nucléotide A au niveau du site de branchement. L'épissage s'effectue en deux étapes de trans-estérification. Dans la première étape, le groupe 2'-hydroxyle du résidu A au niveau du site de branchement attaque le phosphate au niveau du GU du site 5' d'épissage. Ceci conduit au clivage de l'exon en 5 'et à la formation d'un lasso intermédiaire. Dans l'étape suivante, une seconde réaction de trans-estérification, qui implique le phosphate (p) à l'extrémité 3' de l'intron et le groupe 3'-hydroxyle de l'exon détaché, lie les deux exons. Cette réaction libère l'intron, toujours sous la forme d'un lasso.

3. Le spliceosome

Le spliceosome est un complexe composé de cinq particules ribonucléoprotéiques (composées d'ARN et de protéines) principales appelées snRNP (*small nuclear ribonucleoproteins*), de plusieurs mégadaltons et localisé dans le noyau des cellules. Cette macromolécule est l'une des plus complexes des cellules eucaryotes, composé au total de 150 à 300 protéines (Rappsilber et al. 2002; Zhou et al. 2002; Wahl et al. 2009). Sa composition est très dynamique, montrant un échange remarquable de protéines d'une étape à l'autre, accompagné d'un remodelage des RNPsn (Wahl et al. 2009). Chaque RNPsn est elle-même composée d'un ou deux petit(s) ARN non codant(s) snRNA (*small nuclear RNA*) ainsi que d'un nombre variable de protéines spécifiques (Wahl et al. 2009). Ces molécules sont indispensables pour les interactions ARN-ARN lors de l'épissage, mais aussi pour la formation, les réarrangements et la dissociation des complexes par des interactions ARN-protéine et protéine-protéine (Staley and Guthrie 1998). Deux types de spliceosomes coexistent dans les cellules eucaryotes et dans les mêmes gènes. Le spliceosome majeur, ou U2-dépendant, est responsable de l'épissage des introns majeurs. Le spliceosome mineur, ou U12-dépendant, est lui responsable de l'épissage des introns mineurs.

a. Le spliceosome majeur

Le spliceosome majeur, responsable de l'épissage de 99% des introns, est composé de 5 RNPsn : U1, U2, U4/U6 et U5. Chaque RNPsn est constituée d'un ARN (ARNsn), d'un nombre variable de protéines spécifiques et, dans le cas des RNPsn U1, U2, U4 et U5, de sept protéines « Sm » (Wahl et al. 2009). Parmi ces protéines spécifiques on trouve des protéines

de liaison à l'ARN (*RBP, RNA binding protein*) telles que les facteurs U2AF (*U2 Auxiliary Factor 2*), SF1 (*splicing factor 1*) et mBBP (*branch point binding protein*), des enzymes telles que des ARN hélicases, des kinases et des phosphatases (Wahl et al. 2009). L'assemblage des complexes RNPsn du spliceosome sur l'ARN pré-messager pour les réactions d'épissage se fait progressivement, en plusieurs étapes successives (Wahl et al. 2009), car aucune des molécules RNPsn n'est capable à elle seule d'assurer les réactions de trans-estérification, faute de centre catalytique suffisamment performant.

Deux voies d'assemblage alternatives existent dès les premières étapes de l'épissage chez les eucaryotes. La plupart des pré-ARNm de mammifères contiennent plusieurs introns dont les tailles varient de plusieurs centaines à plusieurs milliers de nucléotides (Michael and Manyuan 1999), alors que leurs exons ont une longueur plutôt fixe de seulement 120 nucléotides en moyenne (Lander et al. 2001). Lorsque la longueur de l'intron dépasse 250 nucléotides, ce qui est le cas pour la majorité des introns chez les eucaryotes supérieurs et l'homme en particulier, les complexes d'épissage se fixent d'abord au niveau de l'exon (Fox-Walsh et al. 2005), selon un processus appelé « définition de l'exon » (Berget 1995) (**Figure 9**). Lorsque la longueur de l'intron est inférieure à 250 nucléotides, les complexes d'épissages se fixent au niveau de l'intron, selon le modèle « définition de l'intron » (**Figure 9**).

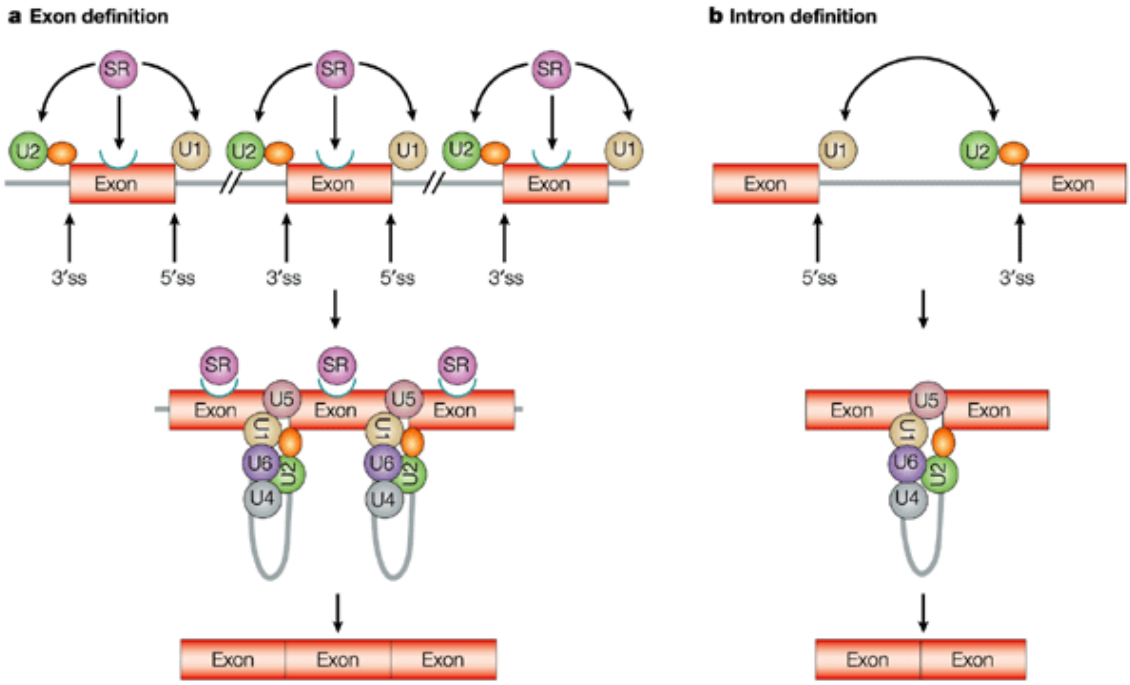

Figure 9 : Complexes dits de définition d'exon (a) et de définition d'intron (b) (d'après Ast 2004). a. Définition de l'exon : les protéines SR (violet) recrutent U1 en aval et le facteur d'épissage U2AF (orange) vers la région riche en pyrimidines en amont. U2AF recrute alors U2 sur le site de branchement. Par conséquent les protéines SR favorisent la formation d'un complexe au niveau de l'exon, en plaçant la machinerie d'épissage au niveau des sites d'épissage du même exon. **b.** Définition de l'intron : la liaison d'U1 au site d'épissage en amont et d'U2AF et U2 au niveau de la région riche en pyrimidines et du site de branchement en aval, se fait au niveau du même intron. Par conséquent, la définition d'intron sélectionne les sites d'épissage situés aux deux extrémités du même intron.

La première étape du recrutement du spliceosome correspond à la liaison de la RNPsn U1 avec le site donneur 5' d'épissage. Cette liaison se fait par appariement de quelques nucléotides (4 à 7) entre l'ARNsn U1 et le site 5' donneur. La liaison est stabilisée par les différentes protéines de la RNPsn U1 ainsi que par d'autres protéines (Puig et al. 1999; Zhang and Rosbash 1999). Lors de cette même étape, les protéines SF1/mBBP et U2AF se lient respectivement au site de branchement et à la séquence poly-pyrimidique (**Figure 9**). Ici, U2AF, protéine composée de deux sous-unités, a sa sous-unité 65 (U2AF65) qui se lie à la protéine SF1/BBP, tandis que sa sous-unité 35 (U2AF35) se lie au site 3' d'épissage. Cet

ensemble, qui fait intervenir les 3 sites d'épissage, constitue le premier complexe du spliceosome : le complexe E (**Figure 10**).

Ces interactions moléculaires déclenchent alors le recrutement de la RNPsn U2 ATP-dépendante au niveau du site de branchement. Ici aussi, l'appariement entre la RNPsn U2 et l'ARN pré-messager est stabilisé par l'action de protéines, notamment SF3a et SF3b (Gozani et al. 1996; Valcárcel et al. 1996). Lors de cette étape, la RNPsn U2 se lie au site de branchement et chasse la SF1/BBP par une réaction ATP-dépendante pour former le complexe A (**Figure 10**). Suite à cette étape, les RNPsn U4/U6 et U5, préassemblées, sont recrutées au niveau du pré-spliceosome pour former le complexe B (**Figure 10**). A ce stade, et bien que toutes les RNPsn aient été recrutées sur l'ARN pré-messager, le complexe B demeure inactif, incapable d'accomplir les réactions de trans-estérification. Ce n'est que lorsque les RNPsn U1 et U4 quittent le complexe que ce dernier subit un réarrangement conformationel lui permettant de devenir catalytiquement actif et formant alors le complexe B* (**Figure 10**). Suite à cela, le spliceosome effectue la première réaction de trans-estérification générant le complexe C, puis subit un autre réarrangement pour en effectuer la deuxième trans-estérification (Konarska et al. 2006). Des ARN hélicases permettent de faciliter ces changements conformationnels de l'ARN (en rose dans la **Figure 10**). Une fois les deux réactions opérées, le spliceosome se dissocie, libérant l'ARN messager et les différents RNPsn, qui pourront être recyclés et être réutilisés pour d'autres réactions d'épissage (**Figure 10**).

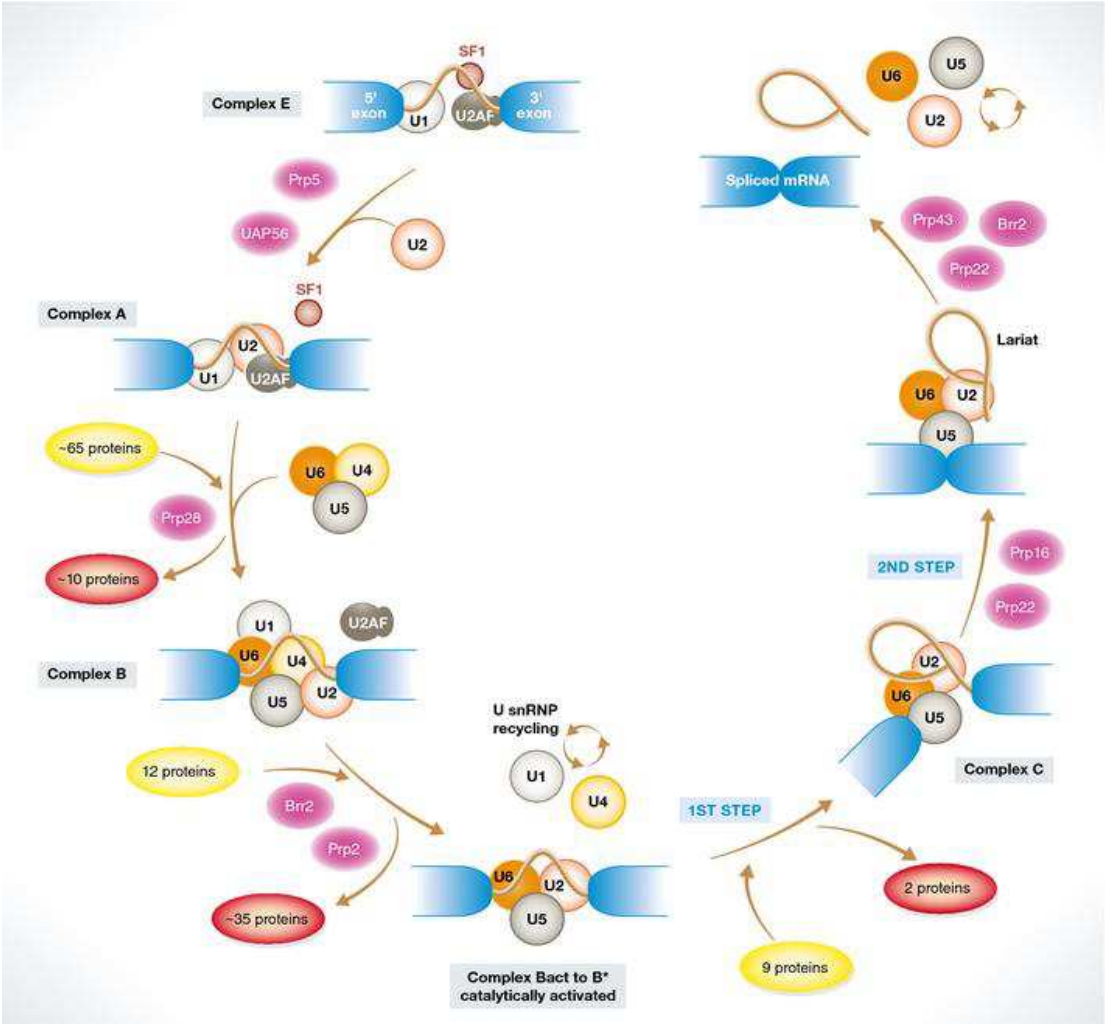

Figure 10 : Les étapes d'assemblage du spliceosome majeur (d'après Daguenet et al. 2015). La reconnaissance initiale du site 5' par la RNPsn U1 et de la région 3' par SF1 (protéine de liaison du site de branchement) et U2AF (complexe E) est suivie par le recrutement du RNPsn U2 au niveau du site de branchement (complexe A), concomitamment avec le déplacement de SF1. La liaison du tri-RNPsn U4/5/6 conduit à la formation du complexe B, impliquant également le déplacement de protéines présentes dans le complexe A. Un remodelage du complexe B, ainsi que le déplacement des RNPsn U1 et U4, conduit à un complexe catalytiquement actif (B*) qui, suite à d'autres réarrangements conformationnels et à des changements dans la composition protéique, catalyse la première étape de la réaction d'épissage, conduisant à la formation d'un lasso intermédiaire (complexe C). Une nouvelle modification conformationnelle conduit à la deuxième étape catalytique, donnant le produit épissé et l'intron en lasso. Lors de la libération des produits, le lasso est linéarisé et dégradé et les RNPsn sont recyclés pour un autre cycle d'assemblage et de catalyse dans d'autres introns.

b. Le spliceosome mineur

Le spliceosome mineur, responsable de l'épissage de 1% des introns, est aussi composé de 5 RNPsn. Quatre de ces RNPsn (U11, U12, U4atac et U6atac) sont différentes mais fonctionnellement analogues à U1, U2, U4 et U6 respectivement. La RNPsn U5 est par contre partagée entre les 2 spliceosomes. Les RNPsn U4atac, U6atac et U5 s'associent préalablement aussi sous forme de trimère (similaire à U4/U6 et U5 dans le spliceosome majeur), alors que les RNPsn U11 et U12 s'associent préalablement sous forme de dimère (contrairement à U1 et U2 dans le spliceosome majeur) (Wassarman and Steitz 1992). L'assemblage global du spliceosome mineur est similaire à celui du spliceosome majeur, la principale différence étant l'absence d'engagement distinct des 2 premières RNPsn, donc l'absence de complexe E (**Figure 11**). Au lieu de cela, le dimère préformé U11/U12 se lie à l'intron, reconnaissant le site 5' d'épissage (*via* U11) en premier puis le site de branchement (*via* U12) sous forme de complexe A (Frilander and Steitz 1999) (**Figure 11**). Comme les introns mineurs n'ont pas de séquence riche en pyrimidines, U2AF n'est pas nécessaire pour leur reconnaissance (Shen and Green 2007). Pour le reste la formation du spliceosome catalytiquement actif est similaire à celle du spliceosome majeur. Le trimère U4atac/U6atac.U5 s'associe au pré-spliceosome pour former le complexe B (Tarn and Steitz 1996). Un grand nombre de réarrangements structuraux convertissent ce complexe B en complexe B* puis C catalytiquement actifs, de manière similaire au spliceosome majeur (**Figure 11**). La RNPsn U6atac remplace U11 au niveau du site 5' d'épissage, puis la RNPsn U4atac est libérée du complexe. Il n'existe pratiquement pas de données montrant que les hélicases utilisées dans le spliceosome majeur aient des activités similaires dans le spliceosome mineur. Cependant, étant donné que la plupart des protéines sont partagées (Schneider et al. 2002; Will et al. 2004) et qu'aucune hélicase spécifique au spliceosome

mineur n'a été identifiée, il est probable que ce soient les mêmes que celles utilisées dans le spliceosome majeur (Turunen et al. 2013).

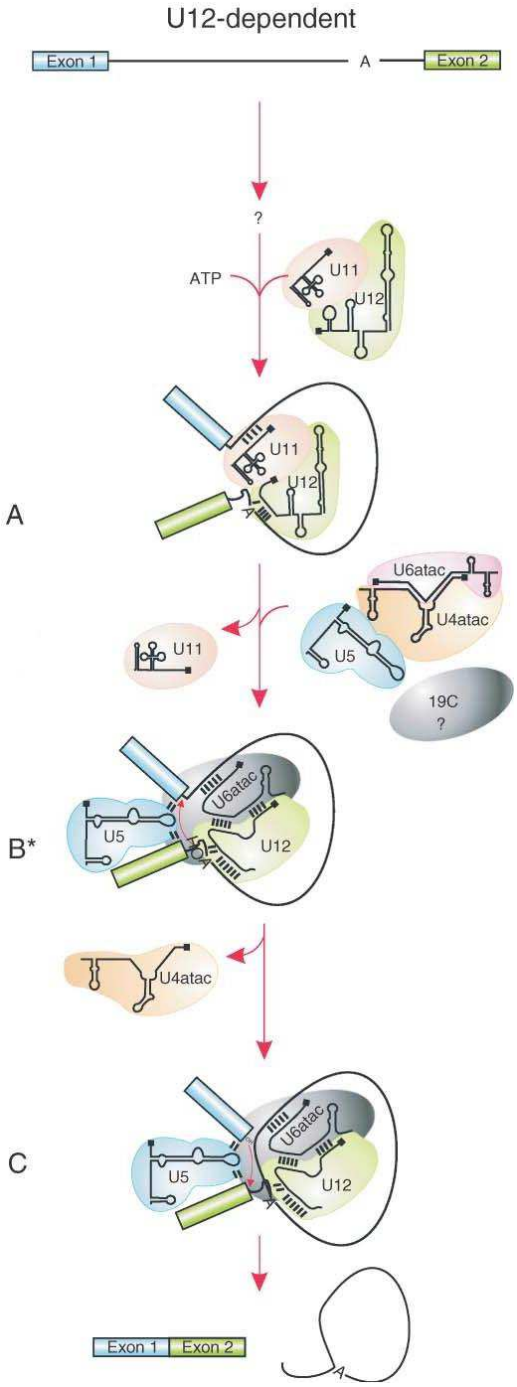

Figure 11 : Etapes d'assemblage du spliceosome mineur (d'après Patel and Steitz 2003). Les étapes d'assemblage sont semblables à celle du spliceosome majeur, à ceci près qu'il n'existe pas de complexe E est que U2AF n'intervient pas dans la formation du complexe.

4. Régulation de l'épissage

De petites séquences, situées dans les exons et les introns, sont directement impliquées dans la régulation de la réaction d'épissage (Fu 1995; Krecic and Swanson 1999). Ces éléments régulateurs d'épissage (*splicing regulatory elements* ou SRE) sont des séquences reconnues par des facteurs spécifiques qui facilitent ou empêchent la reconnaissance des sites d'épissage à proximité (**Figure 12**). Ces protéines peuvent être divisées en 2 grandes classes : les ribonucléoprotéines nucléaires hétérogènes ou hnRNP (*heterogeneous nuclear ribonucleoproteins*) qui ont tendance à inhiber l'épissage et les protéines sérine/arginine (SR) qui ont tendance à activer l'épissage (**Figure 9a**). Alors que les sites consensus d'épissage sont bien connus, les éléments régulateurs d'épissage sont beaucoup moins bien caractérisés.

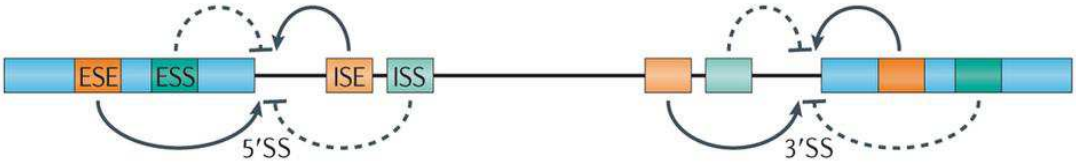

Figure 12 : Les différents éléments régulateurs de l'épissage (d'après Fu and Ares 2014). Les éléments en orange ESE et ISE sont activateurs de l'épissage (flèches pleines) tandis que les éléments en vert ESS et ISS sont inhibiteurs de l'épissage (flèches en pointillés). Abréviations : ESE, élément exonique activateur de l'épissage ; ESS, élément exonique inhibiteur de l'épissage ; ISE, élément intronique activateur de l'épissage ; ISS, élément intronique inhibiteur de l'épissage ; 5'SS, site 5' d'épissage ; 3'SS, site 3' d'épissage.

Longues d'environ 6 nucléotides, les séquences régulatrices d'épissage sont localisées dans tout le pré-ARNm et classées en fonction de leur rôle activateur ou inhibiteur d'épissage (Cartegni et al. 2002; Wang and Burge 2008). Les éléments activateurs d'épissage exoniques, appelés ESE (*exonic splicing enhancers*) et les éléments activateurs d'épissage introniques,

nommés ISE (*intronic splicing enhancers*), sont le plus souvent des séquences riches en purine qui lient des facteurs activateurs d'épissage tels que les protéines SR (**Figures 9a** et **12**). Les éléments inhibiteurs d'épissage exoniques, appelés ESS (*exonic splicing silencers*) et les éléments inhibiteurs d'épissage introniques, nommés ISS (*intronic splicing silencers*), lient des facteurs inhibiteurs d'épissage tels que les protéines hnRNP (**Figure 12**). Les éléments régulateurs de l'épissage les mieux caractérisés sont les éléments exoniques ESE et ESS (Fairbrother et al. 2002; Wang et al. 2006), à l'inverse des éléments introniques ISE et ISS qui sont moins bien connus. Il a été montré que les éléments éloignés des sites d'épissage, aussi bien exoniques qu'introniques, pouvaient être aussi importants que ceux situés à proximité des sites d'épissage (200-300 nucléotides autour du site d'épissage) (Fu and Ares 2014).

a. Eléments activateurs de l'épissage : ESE, ISE et protéines SR

La localisation de ces motifs ESE et ESS est d'une grande importance pour le contrôle de l'épissage, en particulier pour les motifs introniques situés autour des exons. Par exemple, les ESE ont tendance à être regroupés autour des sites d'épissage (Berget 1995) et sont particulièrement présents dans les exons constitutifs, avec des distributions différentes selon les motifs de reconnaissance aux protéines SR (Wang et al. 2005). La grande majorité des exons contient des ESE potentiels (Fairbrother et al. 2002). Les ESE sont reconnus par les protéines SR, appartenant à une famille de protéines apparentées qui partagent un domaine conservé. Elles possèdent une ou deux copies d'un motif de reconnaissance de l'ARN, appelé RRM (*RNA recognition motif*), suivi d'un domaine C-terminal fortement enrichi en dipeptides

arginine/serine (domaine RS), permettant de se fixer à leur substrat de manière spécifique (Birney et al. 1993). Le domaine RS est impliqué principalement, lors de l'assemblage du spliceosome, dans des interactions protéine-protéine, mais apparemment aussi dans des interactions ARN-protéine (Shen and Green 2004; Shen et al. 2004). Il a été décrit 12 protéines SR différentes, nommées de SRSF1 à SRSF12 (Manley and Krainer 2010). Lors de leur synthèse, les protéines SR sont regroupées et concentrées au niveau du noyau avant d'être transportées vers les sites où la transcription est active (Zhou and Fu 2013). Des protéines kinases SRPK (*SR protein kinases*) régulent par phosphorylation la localisation et l'activité de ces protéines SR (Gui et al. 1994). Les protéines SR interviennent également dans le transport de l'ARNm nucléaire (Jensen et al. 2009). Les éléments introniques ISE sont moins bien caractérisés que les éléments exoniques, mais un élément bien connu est le triplet répété GGG (ou *G-run* ; GGG_n avec $n \geq 3$). Souvent distribué en tandem (*clusters*), cet ISE permet de favoriser la reconnaissance du site d'épissage situé à proximité (McCullough and Berget 1997, 2000). Cet ISE est souvent retrouvé dans les introns riches en GC et est conservé entre l'homme et la souris (Yeo and Burge 2004).

b. Eléments inhibiteurs de l'épissage : ESS, ISS et protéines hnRNP

Les éléments ESS et ISS sont reconnus essentiellement par les protéines hnRNP, un groupe divers de protéines exprimées de façon ubiquitaire, dont au moins 20 ont été identifiées, allant de hnRNP A1 à hnRNP U (Dreyfuss et al. 2002). Elles contiennent un ou plusieurs domaines de fixation à l'ARN en N-terminal, appelé RBD (*RNA binding domain*) et pour certaines, comme hnRNP A1, des domaines d'inhibition d'épissage comme des motifs riches en glycine en C-terminal (Dreyfuss et al. 2002; Pozzoli and Sironi 2005). Bien qu'il

reste beaucoup à apprendre sur le fonctionnement précis de ces facteurs, ils peuvent être considérés comme réprimant généralement l'épissage par encombrement stérique ou en favorisant la formation de structures secondaires d'ARN inhibitrices (De Conti et al. 2013). Par exemple, la protéine hnRNP I ou PTB (*polypyrimidine tract binding protein*) se lie au niveau de la région riche en pyrimidine et empêche alors la reconnaissance du site par la protéine U2AF (Izquierdo et al. 2005). PTB peut aussi inhiber l'assemblage du spliceosome par sa liaison au facteur RNPsn U1. La protéine hnRNP A, quant à elle, agit comme antagoniste des protéines SR (Jensen et al. 2009).

5. Epissage alternatif de l'ARNm

L'épissage alternatif est le moyen par lequel deux ou plusieurs ARNm différents peuvent être produits à partir d'un seul gène, contribuant à la diversification de l'expression génique (Graveley 2001). Ainsi, un seul gène peut produire différentes protéines ayant des fonctions identiques, différentes voire opposées. Par exemple, le transcrit pleine longueur du gène *Bcl-x* produit une protéine jouant un rôle dans l'activation de l'apoptose, tandis qu'il existe un transcrit alternatif, plus court, de ce même gène qui produit une protéine inhibant l'apoptose (Boise et al. 1993). La grande majorité des gènes sont soumis à un épissage alternatif (Lee and Rio 2015), un gène humain contenant en moyenne 8 exons et produisant au moins 3 transcrits alternatifs. Alors que certains exons sont présents dans chaque molécule d'ARNm produit, beaucoup sont alternatifs, surtout chez les eucaryotes supérieurs (Wahl et al. 2009). L'épissage alternatif a été décrit comme étant un évènement tissu-spécifique (Heinzen et al. 2008), dépendant du stade de développement (Lee and Rio 2015) et joue un rôle très important dans la différenciation cellulaire (Barbosa-Morais et al. 2012).

a. Différents types d'épissage alternatif

L'épissage alternatif augmente significativement la diversité des transcrits grâce à l'utilisation alternative d'exons ou d'introns, mais aussi de signaux de polyadénylation situés dans des exons alternatifs. Les différents épissages alternatifs peuvent être classés en 5 grands modèles différents (**Figure 13**) :

- Le saut d'un ou de plusieurs exon(s), parfois qualifiés d'exons « cassette » : un ou plusieurs exon(s) entier(s) est (sont) entièrement inclus(s) ou exclu(s) du transcrit mature. C'est le cas d'épissage alternatif le plus simple, qui prédomine chez les mammifères (30 à 50%), donnant lieu à des transcrits matures plus ou moins longs (Kim et al. 2007) ;
- L'utilisation de sites 3' d'épissage alternatif ;
- L'utilisation de sites 5' d'épissage alternatif ;
- Le saut d'exons alternatifs mutuellement exclusifs : les exons en cause ne sont jamais présents ensemble dans le même transcrit mature ;
- La rétention d'intron : l'intron entier est inclus dans le transcrit mature. C'est un évènement rare, qui survient au niveau de petits introns possédant des sites 5' et 3' d'épissage plus faibles (Sakabe and de Souza 2007) ;

Il existe des événements plus complexes faisant intervenir des sites d'initiation de la transcription alternatifs, des sites de polyadénylation alternatifs, ou encore des combinaisons d'événements précités.

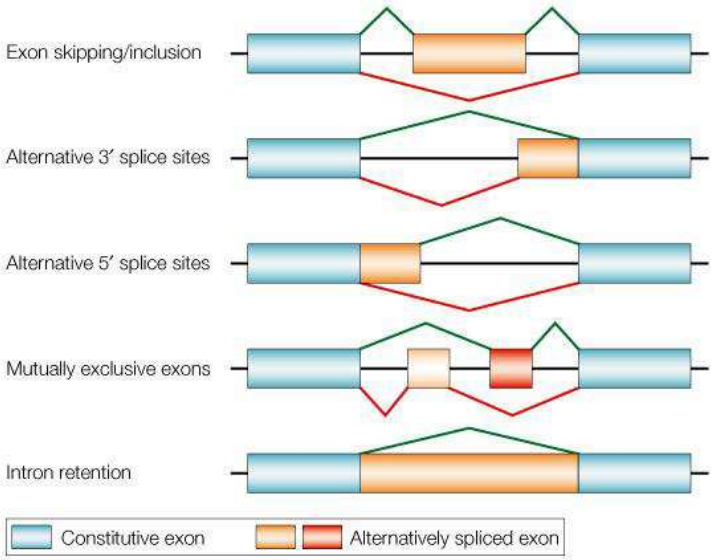

Figure 13 : Différents types d'épissage alternatif (d'après Cartegni et al. 2002a).

La majorité de ces évènements d'épissage alternatif maintient le cadre de lecture, résultant en l'expression de différentes isoformes protéiques. Ces protéines modifiées peuvent être des protéines fonctionnelles, non fonctionnelles, ou avec un effet dominant négatif (Nagy and Maquat 1998). A l'inverse, certains épissages alternatifs engendrent un décalage du cadre de lecture et provoquent l'introduction d'un codon STOP prématuré (PTC, *premature termination codon*) dans le transcrit d'intérêt. Cette rupture du cadre de lecture conduit le plus souvent à la dégradation du transcrit par le système NMD (*nonsense mediated mRNA decay*) (Kugler et al. 1995). Le NMD est un mécanisme de surveillance de l'ARNm chez les organismes eucaryotes qui aide à maintenir la qualité de l'expression génique en accélérant la dégradation des ARNm aberrants porteurs de PTC (Kurosaki and Maquat 2016). Ce mécanisme est activé pour dégrader les transcrits contenant des erreurs et produits pendant le cycle cellulaire, mais aussi pour réguler l'expression des transcrits des gènes en fonction des besoins de la cellule. Pour la dégradation de l'ARNm, 2 voies sont activées : le clivage de la coiffe par une exonucléase et la dégradation de l'ARNm par l'exosome après désadénylation

(Camier and Séraphin 2007). A chaque jonction exon-exon d'épissage, un complexe EJC (*exon junction complex*) se fixe à 20-24 nucléotides en amont de la jonction, jouant le rôle de marqueur de qualité des ARNm. Le codon STOP étant très majoritairement localisé dans le dernier exon, la présence d'un PTC en amont d'une jonction exon-exon marqué par un complexe EJC suggère que ce codon serait précoce (Camier and Séraphin 2007). De fait, la présence d'un complexe EJC en aval d'un PTC est perçue comme un signal de dégradation de l'ARNm correspondant. Le NMD est donc susceptible de dégrader les transcrits possédant un PTC situé à plus de 50-55 nucléotides en amont de la jonction exon-exon (Kurosaki and Maquat 2016).

b. Régulation de l'épissage alternatif

Lorsque l'on considère la régulation de l'épissage, il est utile de distinguer deux classes d'événements d'épissage (Fu and Ares 2014) :

- Des événements d'épissage constitutifs qui sont reconnus efficacement par le spliceosome et qui sont épissés de la même manière dans chaque pré-mRNA à partir d'un gène donné ;
- Des événements d'épissage alternatifs, dans lesquels la reconnaissance et l'appariement entre un site 5' d'épissage et un site 3' d'épissage sont en concurrence avec au moins un autre site 5 'ou 3' d'épissage.

Les sites d'épissage alternatifs ont en commun le fait d'être moins forts que les sites d'épissage constitutifs (Fu and Ares 2014). En revanche, conformément à ce qui est attendu étant donné leur rôle régulateur important conservé dans l'évolution, les sites d'épissage

alternatifs ont tendance à être plus conservés que les sites constitutifs (Sugnet et al. 2004). En plus de la force des sites d'épissage, le choix d'un site d'épissage par le spliceosome est régulé par les SRE (Wang et al. 2006). En effet, les éléments régulateurs d'épissage exoniques ou ESR (*exonic splicing regulator*) sont présents dans tous les exons, constitutifs et alternatifs, mais ils sont plus conservés dans ces derniers (Goren et al. 2006), ce qui explique leur intervention plus fréquente dans l'épissage alternatif. De plus, le même élément ESR peut avoir des effets opposés sur l'épissage lorsqu'il est situé à des positions distinctes dans le même exon alternatif, ce qui suggère que les rôles de régulation de l'épissage par les protéines SR sont hautement position- et contexte-dépendants (Howard and Sanford 2015).

De la même manière, les éléments régulateurs d'épissage introniques ou ISR (*intronic splicing regulator*) situés autour des exons alternatifs sont plus conservés que ceux entourant les exons constitutifs (Jensen et al. 2009), expliquant également leur intervention plus fréquente dans l'épissage alternatif. La raison de cette plus grande conservation s'explique par le fait que les régions introniques proches des exons alternatifs (jusqu'à 150 nucléotides de part et d'autre) sont beaucoup plus conservées que les régions introniques proches des exons constitutifs (Sorek and Ast 2003). Dans ces régions hautement conservées, le motif UGCAUG est particulièrement représenté, avec une conservation depuis les nématodes jusqu'aux mammifères (Minovitsky et al. 2005).

II. Epissages alternatifs dans les gènes *BRCA1* et *BRCA2*

Des études exhaustives des épissages alternatifs dans les gènes *BRCA1* (Colombo et al. 2014) et *BRCA2* (Fackenthal et al. 2016) ont récemment été publiées. Elles contribuent à la compréhension de l'altération de l'épissage de ces gènes dans la prédisposition héréditaires aux cancers du sein et/ou de l'ovaire.

1. *BRCA1*

Le transcrit de référence (transcrit canonique) de *BRCA1* est constitué de 23 exons, dont 22 exons codants. A ce jour, 63 événements d'épissage alternatif physiologiques du gène *BRCA1* ont été identifiés (Colombo et al. 2014). Sur ces 63 événements, 61 ont été détectés dans des lignées lymphoblastoïdes (LLB), 53 dans des cultures primaires de leucocytes du sang périphérique stimulés, 51 dans des leucocytes du sang total, 46 dans des cellules mononucléées du sang périphérique isolées par sédimentation sur ficoll. Trente-neuf événements (62%) sont retrouvés dans les 4 sources d'ARN et 51 événements (81%) dans au moins 3 des sources. Les divergences sont pour la plupart expliquées par la très faible expression de certains transcrits. L'ensemble de ces événements a été rassemblé en 6 groupes selon leur structure : saut d'exon/rétention d'exon cryptique (exon cassette), saut de plusieurs exons (multi-cassette), utilisation d'un site donneur (*distal*, q) ou accepteur (*proximal*, p) d'épissage alternatif, modification terminale et intronisation. L'intronisation étant un évènement rare chez les mammifères (Mudge et al. 2011), seulement deux cas ont été identifiés dans *BRCA1*, dans le très grand exon 11. Il existe des combinaisons d'événements d'épissage, combinant par exemple l'utilisation d'un site donneur alternatif avec le saut de l'exon consécutif. Le transcrit canonique de *BRCA1* (NM_007294.3) et les transcrits

alternatifs en phase qui ont été identifiés dans un premier temps sont décrits dans la **Figure 14**.

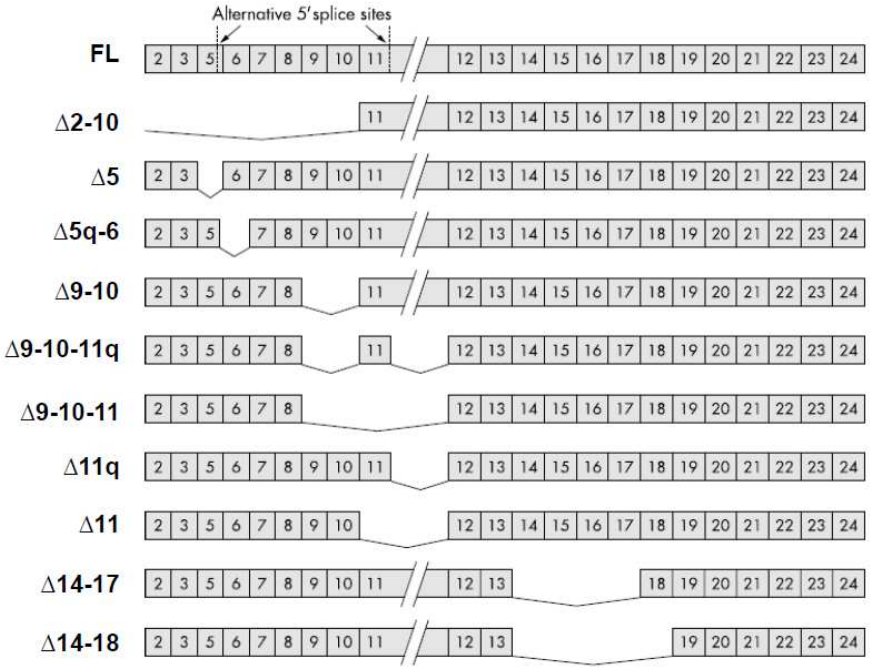

Figure 14 : Différents transcrits alternatifs en phase identifiés dans le gène BRCA1 (d'après Orban and Olah 2003).

Le type d'évènement d'épissage le plus fréquent dans le gène *BRCA1* est le saut d'un ou de plusieurs exons (cassette ou multi-cassette) (n = 37 soit 58%) (Colombo et al. 2014). Il semblerait d'ailleurs que le saut d'exon (30-50% de l'ensemble des épissages alternatifs) soit le type d'épissage alternatif le plus commun observé chez les mammifères (Blencowe 2006; Mudge et al. 2011; Frankish et al. 2012). De manière intéressante, tous les exons internes du gène *BRCA1* sont absents dans au moins un transcrit alternatif, si bien que, de manière formelle, *BRCA1* n'a pas d'exons constitutifs (Colombo et al. 2014). Néanmoins, à part quelques exceptions (Δ5, Δ9, Δ (9,10), Δ9_11), ils concernent les événements d'épissage les

plus faibles, de telle sorte que la plupart des exons apparaissent plutôt comme « quasi-constitutifs ».

En plus du transcrit canonique, 10 transcrits alternatifs de *BRCA1* sont prédominants (Houdayer et al. 2012; Colombo et al. 2014) :

- Δ1Aq (site donneur alternatif dans l'exon 1, délétion de 6 nt en phase) ;
- Δ5 (saut de l'exon 5, délétion de 78 nt en phase) (**Figure 14**) ;
- Δ5q (site donneur alternatif dans l'exon 5, délétion de 22 nt et introduction d'un codon STOP) ;
- Δ8p (site accepteur alternatif dans l'exon 8, délétion de 3 nt en phase) ;
- Δ9 (saut de l'exon 9, délétion de 46 nt hors phase) ;
- Δ(9,10) (saut combiné des exons 9 et 10, délétion de 123 nt en phase) (**Figure 14**) ;
- Δ9_11 (saut combiné des exons 9 à 11, délétion de 3 549 nt en phase) (**Figure 14**) ;
- Δ11q (site donneur alternatif dans l'exon 11, délétion de 3 309 nt en phase) (**Figure 14**) ;
- Δ13p (site accepteur alternatif dans l'exon 13, délétion de 3 nt en phase) ;
- Δ14p (site accepteur alternatif dans l'exon 14, délétion de 3 nt en phase).

Le transcrit Δ1Aq est le transcrit alternatif le plus exprimé car il a été détecté à des niveaux similaires au transcrit pleine longueur tandis que les transcrits Δ8p, Δ(9,10) et Δ14p sont exprimés dans de nombreux tissus et constituent environ 30% de l'expression du transcrit pleine longueur (Colombo et al. 2014). Les isoformes Δ9_11 et Δ11q sont remarquables car il leur manque plus de la moitié de la séquence codante, suggérant que les protéines codées diffèrent grandement de la protéine native. De manière intéressante, le nombre d'événements de type donneur/accepteur exoniques est faible (9/63) mais les transcrits correspondants représentent la majorité des transcrits prédominants (6/10) (Colombo et al. 2014). De plus, il

existe peu d'événements d'épissage de type donneur/accepteur alternatif introniques. Ces événements ne font d'ailleurs pas partie de ceux donnant lieu aux transcrits prédominants. Enfin il n'a pas été identifié d'événement d'épissage de type « exons mutuellement exclusifs » ou « rétention d'intron » (Blencowe 2006; Frankish et al. 2012).

Les conséquences fonctionnelles au niveau des protéines des événements d'épissage détectés semblent variées (Colombo et al. 2014) :

- La plus grande proportion des transcrits alternatifs de *BRCA1* (23/63) induisent un décalage du cadre de lecture conduisant à l'introduction d'un PTC, et sont potentiellement des cibles du NMD ;
- Quinze transcrits sont en phase ;
- Treize transcrits sont non-codant suite à l'élimination du codon d'initiation de la traduction ;
- Cinq transcrits ont une modification des régions terminales non codantes ;
- Quatre transcrits possèdent un PTC non reconnu par le NMD car présent au niveau du dernier exon ;
- Un transcrit provient de l'utilisation d'un site de polyadénylation interne (BRCA1-IRIS) (ElShamy and Livingston 2004). Ce dernier possède un cadre de lecture allant du codon d'initiation de la traduction dans l'exon 2 jusqu'à l'extrémité de l'exon 11 avec en plus 34 codons issus du début de l'intron 11 et a considérablement compliqué l'interprétation des données antérieures issues de modèles murins (Tammaro et al. 2012).

Il semblerait que l'épissage alternatif de *BRCA1* soit similaire dans le sang et dans le tissu mammaire, puisque les 10 événements les plus fréquents détectés dans le tissu

mammaire sont les mêmes que dans le sang et que 43 événements (68%) ont été détecté dans le tissu mammaire (Colombo et al. 2014). Bien que l'étude ait été faite sur un unique échantillon de tissu mammaire, aucun transcrit spécifique à ce tissu n'a été identifié. D'une manière générale, les rôles physiologiques et anormaux (le cas échéant) des épissages alternatifs du gène *BRCA1* restent à définir.

2. *BRCA2*

Le transcrit de référence NM_000059.3 de *BRCA2* est constitué de 27 exons dont 26 codants. A ce jour, 24 transcrits alternatifs du gène *BRCA2* ont été identifiés chez des individus contrôles (Fackenthal et al. 2016). C'est beaucoup moins que pour le gène *BRCA1*, alors que les 2 gènes ont sensiblement la même taille (81 contre 84 kb) et un nombre d'exons proches (23 contre 27). Sur ces 24 événements, 15 ont été détectés dans des leucocytes du sang total (68%), 17 dans des cellules mononucléées du sang périphérique isolées par sédimentation au ficoll (77%), 18 dans des cultures primaires de leucocytes du sang périphérique stimulés (82%) et la totalité des événements a été détectée dans des LLB. Il a aussi été décrit un exon cryptique dans l'intron 12 du gène, très faiblement exprimé chez les individus contrôles (Anczuków et al. 2012). Certains transcrits alternatifs de *BRCA2* contenant un site cryptique donneur/accepteur d'épissage sont décrits dans la **Figure 15**.

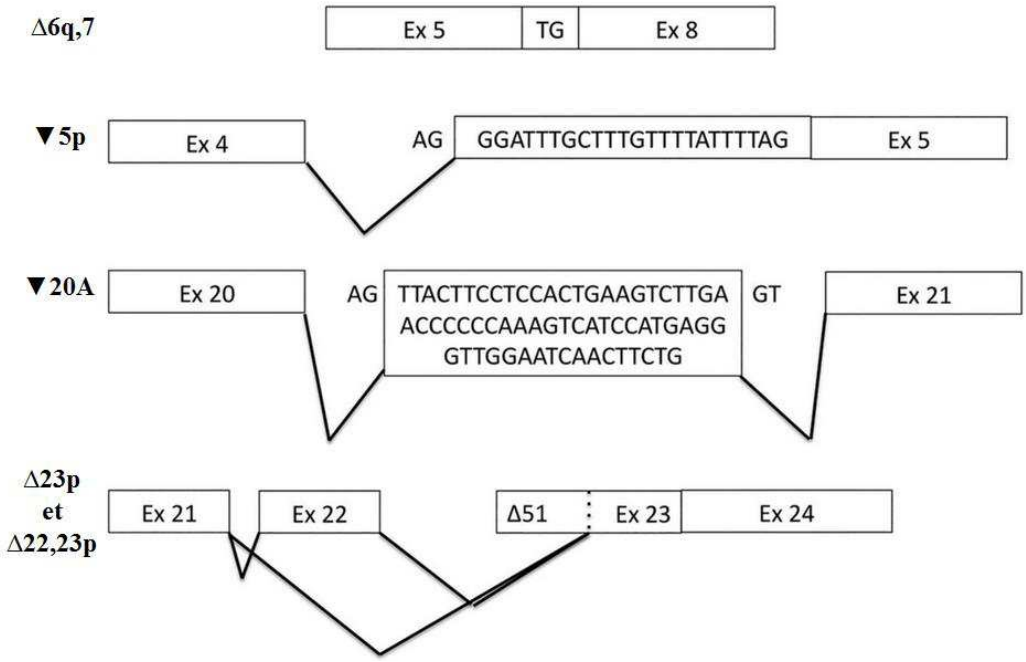

Figure 15 : Structure de certains transcrits alternatifs de *BRCA2* contenant un site cryptique donneur/accepteur d'épissage (d'après Fackenthal et al. 2016). ▼5p correspond à l'utilisation d'un site cryptique accepteur d'épissage situé dans l'intron 4 et retenant les 23 derniers nucléotides de l'intron. ▼20A correspond à l'incorporation d'un exon cryptique situé dans l'intron 20.

Ces transcrits ont été rassemblés en 3 groupes selon leur structure : saut d'exon/rétention d'exon cryptique (exon cassette), saut de plusieurs exons (multi-cassette) et utilisation d'un site accepteur d'épissage alternatif (Fackenthal et al. 2016). Jusqu'à présent il n'a pas été détecté de site donneur d'épissage alternatif, de modification terminale ni d'intronisation. Il existe aussi des combinaisons d'événements d'épissage, combinant par exemple un site donneur alternatif avec le saut de l'exon consécutif. Les évènements les plus fréquemment annotés pour *BRCA2* sont les évènements de type saut d'un ou de plusieurs exons (cassette ou multi-cassette) (n = 20 soit 83%) (Fackenthal et al. 2016). En plus du transcrit canonique, 4 transcrits alternatifs de *BRCA2* sont prédominants :

- Δ3 (saut de l'exon 3, délétion de 249 nt en phase) ;

- Δ6q,7 (combinaison de l'utilisation d'un site donneur alternatif dans l'exon 6, ne gardant que les 2 premiers nucléotides de cet exon, avec le saut de l'exon 7, soit une délétion de 154 nt hors phase) (**Figure 15**) ;
- Δ12 (saut de l'exon 12, délétion de 96 nt en phase) ;
- Δ(17,18) (saut combiné des exons 17 et 18, délétion de 526 nt hors phase) ;

La majorité des transcrits alternatifs de *BRCA2* (18/24) induisent un décalage du cadre de lecture conduisant à l'introduction d'un PTC, et sont potentiellement des cibles du NMD ; 5 transcrits sont en phase et un transcrit est non-codant suite à l'élimination du codon d'initiation de la traduction. Il semblerait aussi que l'épissage alternatif de *BRCA2* soit similaire dans le sang et dans le tissu mammaire, d'après l'étude sur un unique échantillon de tissu mammaire et sur 6 types de lignées cellulaires mammaires différentes (Fackenthal et al. 2016).

III. Anomalies de l'épissage et maladies génétiques

L'épissage de l'ARN dépend donc à la fois d'éléments régulateurs au niveau du pré-mRNA et de facteurs interagissant avec l'ARN mais aussi entre eux (Cooper et al. 2009). Des études pangénomiques ont permis de mettre en évidence que des anomalies du contrôle de l'épissage peuvent être impliquées dans une variété de pathologies à 3 niveaux (Daguenet et al. 2015) :

- Des mutations ou variations génétiques peuvent affecter les éléments régulateurs de l'épissage, nécessaires pour le traitement correct du pré-mRNA, et diminuer la spécificité ou la fidélité de la sélection des sites d'épissage ou activer des sites cryptiques d'épissage normalement peu ou pas utilisés (**Figure 16A**) ;
- Des mutations peuvent affecter les gènes codants les facteurs d'épissage, nécessaires pour la régulation de l'épissage, y compris les composants du spliceosome (**Figure 16B**) ;
- Des profils d'épissage aberrants associés aux maladies génétiques peuvent aussi être dus à un mécanisme d'expansion de répétitions de nucléotides (**Figure 16C**). Dans ce cas, il existe alors un déséquilibre dans la balance des facteurs d'épissage suite à leur séquestration par ces séquences répétées au niveau de l'ARNm, mécanisme aussi appelé « ARN toxique ». Un tel mécanisme peut entrainer des modifications d'expression des gènes à grande échelle.

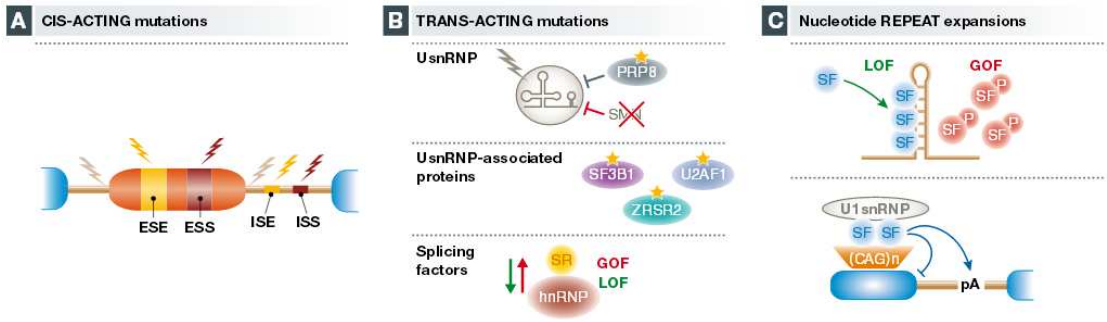

Figure 16 : Mécanismes d'altération de l'épissage dans les maladies génétiques (Daguenet et al. 2015)**. A.** Altérations des éléments régulateurs d'épissage. **B.** Altérations des facteurs d'épissage. **C.** Mécanisme de l'ARN « toxique » par extension de répétitions de nucléotides. SF, splicing factor ; LOF, loss of function ; GOF, gain of function.

Ces altérations de l'épissage peuvent être la cause directe d'une maladie, contribuer à la prédisposition à une maladie ou encore en moduler la sévérité. Elles sont impliquées dans l'étiologie et la progression de nombreuses maladies génétiques monogéniques et multifactorielles telles que les dystrophies musculaires, les syndromes métaboliques, les maladies cardiovasculaires ou neurodégénératives et le cancer (Daguenet et al. 2015). Une altération de l'épissage peut provoquer différents évènements, tels qu'un saut d'exon, une rétention de tout ou partie d'intron ou encore une création/utilisation/destruction d'un site cryptique donneur ou accepteur d'épissage, exonique ou intronique. Les événements d'épissage créant un décalage du cadre de lecture et l'apparition d'un PTC aboutiront le plus souvent à une dégradation rapide du transcrit par le NMD. Si le cadre de lecture est maintenu, ils peuvent alors aboutir à la production d'une protéine tronquée ou modifiée. Enfin, une altération de l'expression de transcrits alternatifs physiologiques peut survenir, modifiant par conséquent les ratios des transcrits, compliquant alors l'interprétation de l'effet des variations sur l'épissage (Ward and Cooper 2010).

1. Altérations des signaux d'épissage

Une variation génétique au sein d'un site d'épissage ou des séquences régulatrices provoquant un épissage aberrant sont souvent retrouvés dans les maladies héréditaires humaines (Daguenet et al. 2015). Des substitutions ponctuelles affectant les sites 5' ou 3' d'épissage sont les mutations d'épissage les plus courantes, entraînant soit un saut d'exon, soit l'activation d'un site cryptique donneur ou accepteur d'épissage, soit dans une moindre mesure une rétention d'intron. De même, des variations introniques ou exoniques (faux-sens, non-sens ou silencieuses) peuvent souvent déclencher des perturbations d'épissage par une perte ou une création d'élément régulateur de l'épissage, activateur ou inhibiteur (Supek et al. 2014; Daguenet et al. 2015). Il est estimé que la moitié des variations synonymes peuvent potentiellement impacter l'épissage (Supek et al. 2014). Ceci est illustré par l'analyse du spectre de mutations du gène *CFTR* (*cystic fibrosis transmembrane conductance regulator*) responsable de la mucoviscidose (Pagani et al. 2005) ou du gène *NF1* responsable de la neurofibromatose, où 50% des mutations délétères conduisent à des défauts d'épissage (Serra et al. 2001). Ces altérations peuvent se produire aussi bien dans les exons constitutifs qu'alternatifs et génèrent par conséquent des transcrits aberrants où il manque un exon constitutif ou entraîner des changements dans les ratios entre isoformes (**Figure 16A**).

Dans la base de données HGMD (*Human Gene Mutation Database*), 9,1% de l'ensemble des mutations publiées sont considérées comme des mutations d'épissage (http://www.hgmd.cf.ac.uk/ac/index.php). Ce pourcentage varie fortement en fonction des gènes : l'on obtient par exemple 23% pour le gène *BRCA2*, 18% pour le gène *ATM* et seulement 9% pour le gène *BRCA1* (Lewandowska 2013). Cependant, ces résultats sont encore très sous-estimés car des mutations potentielles situées dans les éléments régulateurs

pourraient être confirmées expérimentalement. Globalement on estime que les mutations d'épissage représentent un tiers de l'ensemble des mutations délétères (Daguenet et al. 2015). L'utilisation d'algorithmes, tels que MutPred Splice (Mort et al. 2014), a permis de montrer que, parmi les mutations d'épissage identifiées, celles provoquant la perte de sites d'épissage naturels forment la principale catégorie de mutations d'épissage dans les maladies génétiques, dont les prédispositions aux cancers, tandis que celles provoquant la perte d'ESE et/ou le gain d'ESS (générant *in fine* un saut d'exon) forment la principale catégorie de mutations d'épissage dans les cancers sporadiques.

a. Altérations au niveau des sites consensus d'épissage

Une mutation au niveau des sites consensus d'épissage empêcherait *de facto* leur reconnaissance par la machinerie cellulaire que constitue le spliceosome. Les mutations affectant le site de branchement ou la région riche en pyrimidines sont très rares. En revanche, les mutations d'épissage identifiées au niveau des sites 5' et 3' d'épissage sont très fréquentes. Parmi elles, les mutations ponctuelles siégeant au niveau des deux premiers et des deux derniers nucléotides des introns majeurs sont nombreuses (Fredericks et al. 2015) (**Figure 17**). D'autres mutations, à proximité directe de ces dinucléotides, sont fréquemment délétères (Krawczak et al. 2007) (**Figure 17**). Parmi elles figurent :

- Des mutations au niveau du site donneur d'épissage :
    - Au niveau des deux derniers nucléotides exoniques (en position -2 et -1 par rapport au site donneur d'épissage) ;
    - Au niveau intronique en position +3 à +6 ;
- Des mutations du site accepteur d'épissage, en position -3 au niveau de l'intron.

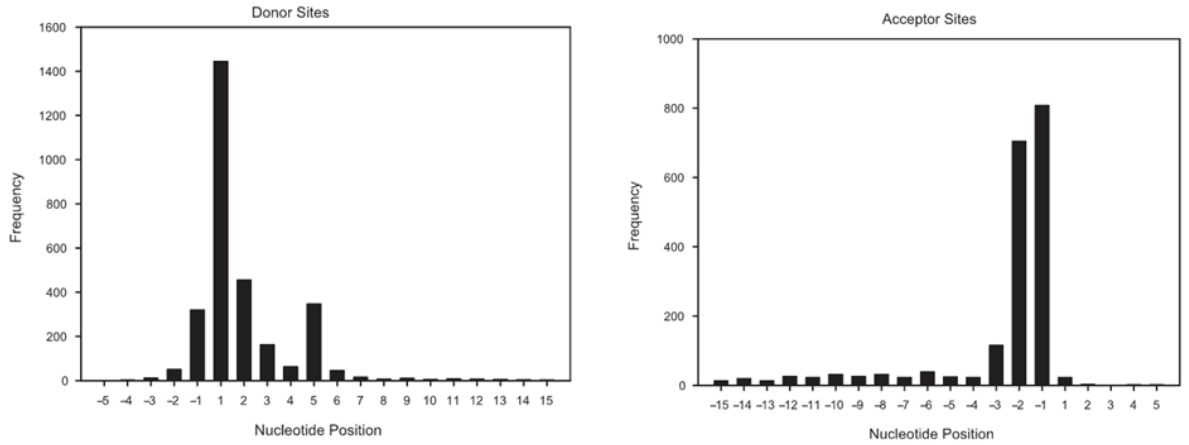

Figure 17 : Distribution des mutations ponctuelles causales répertoriées aux niveaux des sites donneur et accepteur d'épissage dans la base de données HGMD en 2006 (d'après Krawczak et al. 2007)

La destruction d'un site consensus d'épissage empêche donc sa reconnaissance par le spliceosome. En conséquence, celui-ci reconnaitra en principe le site d'épissage accessible le plus proche. Si le site d'épissage choisi est le site consensus de l'exon voisin, cela conduit alors à un saut d'exon. Cependant, le spliceosome peut reconnaitre des sites alternatifs d'épissage, plus proches qu'un autre site consensus. L'utilisation d'un site cryptique peut générer des évènements tels que l'inclusion d'un exon cryptique ou la rétention de l'intron (Krawczak et al. 2007).

Les exemples suivants illustrent comment les mutations pathogènes peuvent être étroitement liées aux multiples aspects de la reconnaissance des sites d'épissage et, en fait, aider à illustrer l'équilibre délicat entre les différents signaux d'épissage et leurs interactions. La grande majorité des patients atteints de dysautonomie familiale (FD, *familial dysautonomia*) porte une mutation ponctuelle à la position +6 de l'intron 20 du gène *IKBKAP* (c.2204+6T>C), qui encode une protéine régulatrice de la transcription (Anderson et al.

2001). Cette position est la dernière des nucléotides introniques du site donneur d'épissage, participant à la reconnaissance du site par la protéine ARNsn U1. La modification de ce nucléotide entraine la perte d'identification du site et donc de la définition de l'exon, provoquant le saut complet de l'exon 20 et générant un ARNm contenant un PTC, cible du NMD (Ibrahim et al. 2007). Le rôle essentiel de ce seul nucléotide est devenu évident lorsque des systèmes expérimentaux ont montré que la restauration de l'appariement de bases entre le site 5' muté et l'ARNsn U1 par mutation de ce dernier restaurait l'inclusion de l'exon 20 (Carmel et al. 2004). De manière remarquable, l'ampleur du défaut d'épissage est tissu-dépendant, très limité dans les lymphoblastes mais très étendu dans le cerveau, expliquant les graves anomalies du cerveau et les symptômes de démyélinisation observés dans la maladie (Cuajungco et al. 2003). Cependant les bases de cette tissu-spécificité restent obscures.

A l'inverse des mutations altérant ou détruisant un site consensus d'épissage, des variations peuvent créer ou activer un site cryptique d'épissage. Ces sites cryptiques, préexistants et situés tout au long des exons et des introns, sont habituellement peu ou pas utilisés par la machinerie d'épissage car leur force est faible comparativement à celle des sites consensus d'épissage. Il n'est donc pas surprenant que de petites variations de séquence peuvent conduire à l'activation de sites cryptiques d'épissage, en augmentant brusquement leur force, comme illustré de façon spectaculaire avec le syndrome de Hutchinson-Guilford (ou progéria) provoquant un vieillissement prématuré et où les individus les plus touchés portent une substitution silencieuse C>T dans l'exon 11 du gène *LMNA* (lamine A) en position c.1824. Cette mutation active un site cryptique 5' d'épissage, générant un ARNm tronqué des 150 derniers nucléotides de l'exon 11 et codant une forme dominante négative de lamine A, tronquée de 50 acides aminés et appelée progérine, causant une instabilité nucléaire et génomique.

Des altérations des principaux signaux d'épissage peuvent aussi être dues à une altération de la région riche en pyrimidines. Ces altérations sont rares mais peuvent être à l'origine de certaines maladies comme l'hémophilie B, maladie récessive liée au chromosome X (Lewandowska 2013). Actuellement, dans cette pathologie, les variations dans le site de branchement et la région riche en pyrimidine sont recherchées quand aucune altération n'est identifiée par séquençage dans la région codante, dans les sites 5' et 3' d'épissage, et dans les régions transcrites non traduites (UTR, *untranslated regions*).

Enfin, des mutations ponctuelles à proximité des sites d'épissage peuvent considérablement affecter l'épissage et provoquer la maladie. Un exemple remarquable est celui des FTDP-17 (*frontotemporal dementia and parkinsonism linked to chromosome 17*), un syndrome démentiel lié à des mutations du gène *MAPT* (*microtubule associated protein Tau*). Des mutations au sein de l'intron 10 du gène, à proximité du site 5' d'épissage, altèrent l'équilibre entre les différentes isoformes d'épissage en augmentant l'inclusion de l'exon 10. L'augmentation d'expression de cet isoforme conduit à l'agrégation de la protéine Tau, responsable d'une dégénérescence neuronale. Cette mort neuronale est à l'origine d'une démence se traduisant initialement par des troubles de la personnalité et associant parfois des troubles moteurs (Hutton et al. 1998). Les mutations ne sont pas situées au niveau des sites canoniques d'épissage AG/GT mais induisent l'ouverture d'une tige-boucle qui normalement séquestre partiellement le site 5' d'épissage afin d'éviter l'inclusion totale de l'exon 10 (Hutton et al. 1998; Grover et al. 1999; Pottier et al. 2016). Ceci est l'un des exemples les mieux documentés sur l'influence des structures secondaires du pré-mRNA sur la reconnaissance du site d'épissage.

b. Altérations au niveau des sites régulateurs de l'épissage

L'altération des séquences régulatrices de l'épissage par des mutations est un mécanisme fréquent et probablement sous-estimé (Soukarieh et al. 2016). En effet, il est estimé que 25% des variations exoniques de type faux-sens et non-sens altèrent l'épissage par création ou destruction d'ESE et d'ESS (Sterne-Weiler et al. 2011). Ces données montrent que toute variation peut potentiellement être une mutation d'épissage, indépendamment de son effet potentiel sur la structure ou la fonction de la protéine. De plus, conformément à l'observation que les gènes liés au cancer peuvent posséder une plus grande sensibilité envers des altérations de l'épissage (Sterne-Weiler and Sanford 2014), des mutations synonymes récurrentes affectant des ESE et ESS ont été identifiées dans des oncogènes importants, représentant un mécanisme supplémentaire dans l'activation de ces derniers (Daguenet et al. 2015).

Un exemple classique de la façon dont une mutation exonique peut influencer la reconnaissance de l'exon avec de graves conséquences est l'amyotrophie spinale (SMA, *spinal muscular atrophy*). C'est l'une des maladies génétiques les plus fréquentes, maladie neuromusculaire autosomique récessive caractérisée par la perte sélective des motoneurones spinaux, conduisant à une faiblesse sévère des muscles squelettiques associée à une atrophie. Elle implique une quantité insuffisante de protéine SMN (*survival of motor neuron*) dont la fonction est de chaperonner la biogénèse et l'assemblage des RNPsn (Matera and Wang 2014). Cette insuffisance quantitative résulte de mutations de type perte de fonction ou de délétions dans le gène *SMN1* (Lefebvre et al. 1995). Malgré sa forte homologie avec *SMN1*, le gène *SMN2*, issu de la duplication génique de *SMN1*, ne parvient pas à empêcher le développement de la maladie car ce gène porte une substitution nucléotidique au niveau du 6^{e}

nucléotide de l'exon 7 (C6T ex7 ou c.840C>T), responsable du saut de l'exon 7 et générant une version tronquée, instable et rapidement dégradée de la protéine SMN (Lorson et al. 1999). Des analyses de l'impact mécanistique de la transition C vers T ont montré la perte d'un ESE reconnu par la protéine SRSF1 en même temps que le gain d'un ESS reconnu par la protéine hnRNPA1 (Cartegni and Krainer 2002; Kashima and Manley 2003). Ce modèle montre le caractère inconstant des signaux de régulation d'épissage, pouvant convertir un ESE en ESS à la suite d'une modification d'un seul nucléotide.

2. Altérations des facteurs régulateurs de l'épissage

Bien qu'il existe un nombre important et croissant d'exemples de mutations d'épissage agissant au niveau des signaux d'épissage et causant des maladies, les exemples de mutations d'épissage agissant au niveau des facteurs d'épissage sont relativement limités (Singh and Cooper 2012). La rareté des mutations agissant au niveau des facteurs d'épissage suggère que ces mutations au sein de la machinerie d'épissage sont létales pendant le développement embryonnaire, sinon au niveau des cellules individuelles (Kalsotra and Cooper 2011). Cependant, des études décrivent des mutations pathogènes agissant au sein de la machinerie d'épissage, y compris au sein même des composants du spliceosome ainsi que d'autres composants nécessaires pour la régulation de l'épissage alternatif (Singh and Cooper 2012) (**Figure 16**).

a. Altérations de la biogénèse des RNPsn

Comme décrit précédemment, la mutation du gène *SMN1* conduit à des défauts dans la biogénèse des RNPsn responsable de la SMA. De manière intéressante, des liens moléculaires potentiels entre la fonction des RNPsn dans la SMA et d'autres troubles moteurs et/ou neurodégénératifs ont émergé. La sclérose latérale amyotrophique (SLA) en est un bel exemple, causée par des mutations dans plus de 20 gènes ayant des fonctions différentes (Renton et al. 2014), y compris les protéines FUS (*fused in sarcoma*) et TDP-43 (*transactive response DNA binding protein 43 kDa*) se liant à l'ARN. FUS est une protéine hnRNP-*like* qui interagit avec les ARNsn U1 et U2 (Kwiatkowski et al. 2009; Vance et al. 2009; Gerbino et al. 2013). Les protéines FUS mutantes se lient également à ces ARNsn, avec une moins bonne affinité, mais sont ensuite retenues dans le cytoplasme, provoquant une diminution du pool disponible de RNPsn U1 et U2 dans le noyau (Yu et al. 2015). En outre, FUS interagit avec SMN et les protéines FUS mutées semblent également modifier la localisation cellulaire de SMN, contribuant potentiellement aussi aux altérations d'épissage associées à la maladie (Yamazaki et al. 2012; Groen et al. 2013; Sun et al. 2015).

La protéine TDP-43, autre protéine se liant à l'ARN et fréquemment mutée dans la SLA, interagit avec FUS (Groen et al. 2013). Son altération entraine en conséquence une dérégulation de FUS et ainsi conduit à une perte d'abondance des ARNsn et la mauvaise localisation de SMN (Ishihara et al. 2013). La protéine TDP-43 mutée forme des agrégats insolubles (Neumann et al. 2006), marqueurs caractéristiques d'autres maladies neurodégénératives, telles que la maladie d'Alzheimer. La récente analyse de la teneur en protéines des agrégats retrouvés dans la maladie d'Alzheimer a montré l'accumulation de certains composants de la RNPsn U1 (Bai et al. 2013) et des analyses immuno-histochimiques

ont révélé que les protéines U1-70K et U1A forment des enchevêtrements dans le cytoplasme des cellules du cerveau de malades, mais pas dans les autres maladies neurodégénératives. De plus, des analyses de séquençage à haut débit de l'ARN (RNA-Seq) ont révélé un défaut global d'épissage, avec l'accumulation d'ARN non épissés (Bai et al. 2013), ce qui suggère un rôle de l'épissage de l'ARN dans l'étiologie de cette maladie. Ces observations montrent donc que des altérations menant à l'accumulation cytoplasmique de RNPsn et par conséquent à des altérations de l'épissage des pré-mRNA contribuent à l'étiologie de multiples maladies neurodégénératives peut-être grâce à des mécanismes et des cibles communes (Daguenet et al. 2015).

Des altérations de la biogénèse de l'ARNsn U6 ont été détectées dans la poïkilodermie avec neutropénie (PN) de type Cléricuzio, rare maladie de peau autosomique récessive fréquemment associée à une neutropénie chronique et des anomalies de la moelle osseuse pouvant aboutir à une myélodysplasie et un risque accru de transformation leucémique (Shchepachev et al. 2012). La maladie est associée à des mutations dans le gène *C16orf57* (appelé aussi *USB1*, *U6 snRNA biogenesis 1*) codant la protéine hMpn1/Usb1 (Volpi et al. 2010). Cette exoribonucléase 3'→5' est impliquée dans la maturation de l'ARNsn U6, supprimant sa queue de résidus uridine et générant un phosphate 2'-3' cyclique qui stabilise l'ARNsn (Shchepachev et al. 2012). La baisse du taux d'enzyme dans les lymphoblastes de patients atteints de PN se traduit par une augmentation de la dégradation de l'ARNsn U6. Cependant, aucune perturbation globale de l'épissage n'a été observée dans ces cellules (Shchepachev et al. 2012), ce qui est remarquable compte tenu du rôle crucial que l'ARNsn U6 joue dans les 2 étapes catalytiques du processus d'épissage (**Figure 10**). Ces résultats suggèrent que la maladie est soit associée à de légers effets spécifiques sur l'épissage de

certains ARN, soit que cette enzyme joue un rôle dans le métabolisme d'autres classes d'ARN (Daguenet et al. 2015).

b. Altérations des protéines composant le spliceosome

Des données récentes de séquençage de génomes et d'exomes de cellules sanguines anormales de patients souffrant de maladies hématopoïétiques des lignées lymphoïdes et myéloïdes, y compris de syndrome myélodysplasique (SMD), de leucémie aigüe myéloïde (LAM), de leucémie lymphocytaire chronique (LLC) et de leucémie myélomonocytaire chronique (LMMC), ont mis en évidence des mutations somatiques récurrentes dans des gènes codants des facteurs d'épissage (Daguenet et al. 2015). Parmi ces facteurs impliqués, on trouve des protéines impliquées dans la reconnaissance du site 3' d'épissage, telles que 2 sous-unités du complexe U2AF, SF1/BBP et ZRSR2 (*zinc finger (CCCH type), RNA-binding motif and serine/arginine rich 2*), et des protéines composant le complexe RNPsn U2, comme SF3A1 (*splicing factor 3a, subunit 1, 120kDa*) et SF3B1 (*splicing factor 3b, subunit 1, 155kDa*) (Papaemmanuil et al. 2011; Yoshida et al. 2011; Quesada et al. 2012). Dans ces données, le facteur le plus fréquemment muté était le gène *SF3B1*, dont la protéine joue un rôle essentiel dans la reconnaissance du site d'épissage. Ses interactions avec U2AF et le pré-mRNA contribuent au recrutement du RNPsn U2 (Gozani et al. 1998) ; sa phosphorylation est couplée avec les réactions de catalyse pendant l'épissage (Wang et al. 1998) et son association à la chromatine influence les réactions d'épissage (Kfir et al. 2015). Des mutations de *SF3B1* ont été identifiées dans 85% des patients souffrant de SMD couplée à une anémie réfractaire avec « sidéroblastes en couronne », 15% des patients atteints de LLC et 5% des patients souffrant de LAM ou de LMMC (Daguenet et al. 2015). Etonnamment, alors que les

mutations de *SF3B1* sont corrélées à un pronostic favorable dans les SMD, elles sont corrélées à un pronostic défavorable couplé à une résistance à la fludarabine dans les LLC (Papaemmanuil et al. 2011; Rossi et al. 2011), suggérant que les effets moléculaires de ces mutations sont fortement influencés par le contexte cellulaire (Daguenet et al. 2015). Ce concept est en outre vérifié par la récente identification d'altérations dans un autre composant du complexe RNPsn U2, SF3B4 (*splicing factor 3b, subunit 4, 49kDa*) chez des patients atteints du syndrome de Nager, dysostose acro-faciale caractérisée par des malformations cranio-faciales et des membres (Bernier et al. 2012). De manière intéressante, des analyses de séquençage à haut débit d'ADN de sang de plus de 4000 individus ont révélé que l'expansion de clones hématopoïétiques portant des mutations dans les gènes *SF3B1* et *SRSF2* augmente fortement avec l'âge, ce qui est compatible avec l'incidence plus élevée des SMD à un âge avancé (McKerrell et al. 2015).

La régulation de l'épissage s'opère au niveau de toutes les étapes de l'assemblage du spliceosome. Il est donc attendu qu'une mutation altérant l'un de ces facteurs soit préjudiciable au moment de l'élimination des introns (Daguenet et al. 2015). Par exemple, des mutations dans les composants du complexe RNPsn U4/U6/U5 conduisent à des syndromes très spécifiques. Ainsi, des mutations dans le gène *EFTUD2* (*elongation factor Tu GTP binding domain containing 2*), qui code pour la protéine U5-116kDa, composant du RNPsn U5, sont responsables de la dysostose mandibulo-faciale avec microcéphalie, syndrome poly-malformatif (Lines et al. 2012). De plus, des mutations dans les gènes *PRPF31*, *PRPF8*, *PRPF6*, *PRPF3* (*pre-mRNA processing factors*), *RP9* (*retinitis pigmentosa (autosomal dominant)*, codant la protéine PAP-1 et SNRNP200/BRR2 (*small nuclear ribonucleoprotein 200kDa*), sont associées aux rétinites pigmentaires, large groupe de maladies héréditaires dégénératives de la rétine caractérisées par un dysfonctionnement progressif des

photorécepteurs et de l'épithélium pigmentaire de l'œil (Mordes et al. 2006). On peut se demander pourquoi des altérations de facteurs clés de la régulation de l'épissage conduisent à des pathologies tissu-spécifiques. Il est possible qu'une altération de l'épissage soit plus préjudiciable dans des cellules à division rapide composant des tissus nécessitant une régénération rapide, comme la rétine (Daguenet et al. 2015).

Bien qu'utilisé pour l'épissage d'un nombre très limité d'introns, le spliceosome mineur peut également être la cible de mutations délétères. Ainsi, des mutations dans le gène codant l'ARNsn U4atac sont responsables du nanisme microcéphalique ostéodysplasique primordial de type I (appelé aussi syndrome de Taybi-Linder), caractérisé par des anomalies neurologiques et squelettiques (Edery et al. 2011; He et al. 2011).

3. Expansion de répétitions de nucléotides – ARN « toxique »

La maladie associée à une expansion de microsatellites la mieux comprise est la dystrophie myotonique (DM). Un nombre anormal de répétitions CTG dans la région du gène *DMPK* (*dystrophia myotonica-protein kinase*) codant pour la 3'UTR (Mahadevan et al. 1992) et un nombre tout aussi anormal de répétition CCTG dans l'intron 1 du gène *CNBP/ZNF9* (*CCHC-type zinc finger, nucleic acid binding protein*) (Liquori et al. 2001), conduisant dans les 2 cas à un gain de fonction, sont responsables respectivement de la DM de type 1 (maladie de Steinert) et de type 2 (maladie de Ricker) (Cooper et al. 2009; Ward and Cooper 2010). Ces grandes répétitions CUG et CCUG au sein des pré-mRNA de ces gènes peuvent se replier sur elles-mêmes et former de relativement longues structures d'ARN double-brin, avec deux conséquences moléculaires distinctes impactant la régulation de l'épissage (**Figure 16C**). D'une part, les pré-ARNm porteurs de répétitions, qui s'accumulent dans des foyers nucléaires, se lient ensemble et séquestrent MBNL1 (*muscleblind-like splicing regulator 1*), un régulateur de l'épissage alternatif important pour la différenciation cellulaire (Lin et al. 2006). D'autre part, les répétitions CUG stabilisent la protéine CUG-BP1 (*CUG-binding protein 1*) par l'intermédiaire d'une phosphorylation accrue médiée par l'activation de PKC (*protein kinase C*) (Kuyumcu-Martinez et al. 2007). La diminution de l'activité de MBNL1 et de l'augmentation de l'activité de CUG-BP1 entrainent des modifications de la régulation de l'épissage alternatif pouvant expliquer les principales caractéristiques de la maladie. Par exemple, la myotonie (retard à la relaxation musculaire) est liée à l'inclusion aberrante d'un exon alternatif décalant le cadre de lecture du gène codant le canal chlore musculaire *CLNC1* (*chloride channel protein, skeletal muscle*) (Charlet-B et al. 2002; Mankodi et al. 2002), tandis que l'insulino-résistance est liée à un saut d'exon dans le pré-mRNA du récepteur à l'insuline conduisant à la production d'un récepteur moins sensible (Savkur et al. 2001).

La dystrophie myotonique peut servir de paradigme pour comprendre d'autres troubles causés par des expansions répétées (Cooper et al. 2009), y compris certaines formes de SLA, l'ataxie/tremblements associés au syndrome de l'X fragile, l'ataxie spinocérébelleuse mais aussi la maladie de Huntington. Cette maladie neurodégénérative, caractérisée par des mouvements involontaires, des symptômes psychiatriques ainsi qu'une démence, est causée par une expansion de répétitions CAG dans l'exon 1 du gène *HTT* codant la huntingtine (MacDonald et al. 1993). Si le nombre de répétitions CAG est supérieur à 40, les symptômes apparaissent, le nombre de répétitions étant corrélé avec la sévérité et la précocité de la maladie. Classiquement, les effets pathologiques de l'expansion ont été attribués à un gain de fonction en raison de l'augmentation du nombre d'acides aminés glutamines (polyglutamines) encodés par la répétition de codons CAG dans la protéine HTT. On observe alors une accumulation de HTT ou de produits de dégradation de la région N-terminale riche en glutamine dans des corps d'inclusion. Une étude récente a rapporté un site de polyadénylation cryptique au sein de l'intron 1 du gène, qui est activé lors de l'expansion des CAG (Sathasivam et al. 2013). Cet ARN est activement traduit en protéine plus courte et contenant des régions polyglutamines, donnant ainsi une autre explication sur la génération de peptides N-terminaux toxiques (Daguenet et al. 2015). Fait intéressant, la protéine SR SRSF6 semble se lier aux répétitions CAG de ces transcrits (car porteurs de sites consensus pour cette protéine) et favorise donc l'inhibition de l'épissage de l'intron 1, ce qui facilite l'utilisation du site de polyadénylation alternatif (Sathasivam et al. 2013). Une hypothèse intéressante est que SRSF6 empêcherait l'association de la RNPsn U1 avec le site 5' d'épissage de l'intron 1, expliquant à la fois l'inhibition de l'épissage et l'activation du site cryptique de polyadénylation (Berg et al. 2012). Un lien intéressant a été récemment montré entre la maladie de Huntington et les tauopathies (Fernández-Nogales et al. 2014). Dans cette étude,

un déséquilibre entre les différents transcrits alternatifs de Tau, semblable à celui retrouvé dans le syndrome FTDP-17 (voir précédemment), a été observé dans le cerveau de patients atteints de la maladie de Huntington, ainsi que des dépôts de protéine Tau dans les noyaux neuronaux contribuant au phénotype moteur chez des souris transgéniques mutées *HTT*. De manière remarquable, SRSF6 est un régulateur connu de l'épissage de l'exon 10 de *Tau* et l'association des protéines SRSF6 avec les transcrits comportant des expansions de CAG peut potentiellement expliquer non seulement la génération de transcrits *HTT* plus courts, mais aussi l'altération de l'épissage de *Tau* (Fernández-Nogales et al. 2014). Par conséquent, les changements d'activité de SRSF6 induits par l'augmentation des répétitions CAG peuvent être la cause de l'épissage aberrant des 2 gènes *HTT* et *Tau*, et des dépôts des 2 protéines HTT et Tau (Daguenet et al. 2015).

4. Implication de l'épissage dans la prédisposition héréditaire aux cancers du sein et de l'ovaire

Les mutations abolissant les sites consensus d'épissage ou les éléments régulateurs de l'épissage sont fréquentes dans les gènes responsables de prédispositions génétiques aux cancers (Mort et al. 2014). La majorité de ces mutations se situe au niveau des sites consensus d'épissage, à la fois au niveau des premiers et derniers dinucléotides introniques invariants mais aussi au niveau des nucléotides introniques et exoniques moins conservés (Venables 2004; Houdayer et al. 2012).

Les substitutions ponctuelles affectant les sites consensus 5' et 3' d'épissage sont fréquentes dans les gènes *BRCA1* et *BRCA2* (Houdayer et al. 2012). Par exemple, la variation c.134+1G>A située dans l'intron 3 du gène *BRCA1* détruit le site 5' d'épissage. De même, la variation c.5278-1G>C située dans l'intron 20 détruit le site 3' d'épissage. En ce qui concerne les nucléotides moins conservés, la mutation de la dernière base de l'exon 14 (c.4484G>T) provoque le saut de l'exon 14, celle d'un nucléotide intronique moins conservé dans l'intron 22 (c.5406+5G>C) provoque le saut de l'exon 22, à chaque fois par destruction du site 5' d'épissage.

Dans le gène *BRCA2*, les variations c.67+1G>T (intron 2) et c.426-2A>T (intron 4) touchent les nucléotides invariants des sites consensus d'épissage et sont responsables respectivement du saut de l'exon 2 et de l'exon 5. Tandis que concernant les nucléotides moins conservés, la mutation de la dernière base de l'exon 23 (c.9117G>A) provoque le saut de l'exon 23, celle d'un nucléotide intronique moins conservé dans l'intron 3 (c.316+5G>C) provoque le saut de l'exon 3, à chaque fois par destruction du site 5' d'épissage.

D’autres variations sont susceptibles d’affecter des éléments régulateurs de l’épissage. Par exemple, la mutation c.5434C>G (p.Pro1812Ala) située au milieu du petit exon 23 du gène *BRCA1* (61 nt) provoque le saut dudit exon par destruction d’un site exonique régulateur d’épissage (ESR) (Gaildrat et al. 2010b). Il a aussi été rapporté des exemples de création ou de renforcement de sites 5’ ou 3’ d’épissage cryptiques. Par exemple, la mutation c.6937+594T>G située à distance des sites d’épissage physiologiques, dans l’intron 12 du gène *BRCA2*, renforce un site 5’ d’épissage cryptique, ce qui provoque l’inclusion de l’exon cryptique de 95 nucléotides (Anczuków et al. 2012). Cet exemple est remarquable car les variations introniques profondes ne sont pas ou peu recherchées en diagnostic, le séquençage étant limité aux exons et aux introns proches. De manière plus générale, aujourd’hui, la vision des altérations de l’épissage est probablement biaisée car les publications et bases de données sont surtout enrichies en mutations affectant les sites canoniques d’épissage AG/GT. Il a d'ailleurs récemment été montré que la contribution des mutations de l'épissage est probablement très sous-estimée (Soukarieh et al. 2016).

Des mutations d’épissage ont aussi été détectées dans d’autres gènes impliqués ou suspectés de l’être dans la prédisposition aux cancers du sein et/ou de l’ovaire. En particulier, l’étude des paralogues du gène *RAD51* a permis d’identifier des mutations d’épissage (Golmard et al. 2013). Le variant c.706-2A>G situé sur un nucléotide invariant du site accepteur d’épissage au niveau de l’exon 5 du gène *RAD51C* provoque le saut de cet exon, tandis que la délétion de 3 bases au niveau de nucléotides moins conservés du site donneur d’épissage au niveau de l’exon 8 (c.1026+5_1026+7del) provoque le saut de l’exon 8.

Au total, dans ces cancers héréditaires du sein et/ou de l'ovaire, les nombreuses altérations de l'épissage observées génèrent le plus souvent des protéines tronquées ou

délétées d'un domaine fonctionnel, à l'origine d'une perte de fonction protéique de ces gènes suppresseurs de tumeurs. L'implication physiopathologique du renforcement des épissages physiologiques de ces gènes n'est pas bien connue. Seul l'épissage physiologique *BRCA-IRIS* a été étudié. Il aurait d'ailleurs un rôle potentiel oncogénique (Shimizu et al. 2012).

IV. Analyse des effets sur l'épissage des variations dans les gènes du syndrome HBOC

1. Prédictions bio-informatiques

Sachant que toute variation peut potentiellement être une mutation d'épissage, l'étude expérimentale de l'impact sur l'épissage de la totalité des variations détectées chez les patientes n'est pas, à ce jour, envisageable. L'utilisation d'algorithmes de prédictions de l'impact sur l'épissage d'un VSI permet de prioriser les variations les plus pertinentes à tester (Houdayer et al. 2008; Spurdle et al. 2008; Soukarieh et al. 2016).

Différents algorithmes de prédictions ont été développés pour évaluer la position et estimer la force des sites 5' et 3' d'épissage. Ils calculent d'abord un score évaluant la force de chaque site 5' et 3' d'épissage, qu'il soit consensus ou cryptique. Ils calculent ensuite le même score après apparition de la variation à tester, permettant alors de prédire l'effet potentiel de la variation sur l'épissage, par comparaison des scores. Les variations fortement suspectées d'affecter l'épissage peuvent :

- Diminuer fortement la force d'un site consensus d'épissage, pouvant provoquer le saut d'un ou de plusieurs exon(s) ;
- Augmenter la force et donc l'utilisation d'un site cryptique d'épissage déjà existant à proximité ;
- Créer un nouveau site d'épissage.

Il est aussi possible qu'un variant puisse à la fois diminuer la force d'un site consensus d'épissage, tout en augmentant la force d'un site cryptique ou en créant un nouveau site

d'épissage. En comparant les scores des sites physiologiques ou cryptiques proches, il est possible de prédire l'utilisation préférentielle d'un site par rapport à un autre (Desmet et al. 2009).

Différents outils de prédictions bio-informatiques sont disponibles, et utilisent des algorithmes différents (Spurdle et al. 2008). Parmi les plus utilisés figurent :

- SSF (*Splice Site Finder*), dont les scores ont été établis sur la base des fréquences attribuées aux nucléotides à chaque position des sites d'épissage (Shapiro and Senapathy 1987). Cette méthode prédit les exons potentiels dans une séquence, en utilisant ces scores et un système de classement basé sur des tables de poids nucléotidiques (PWM, *position weight matrix*). Les scores varient de 0 à 100, sachant que plus la valeur du score prédit est grande, plus la probabilité que le site soit un vrai site d'épissage est élevée ;
- NNSplice (*Neural Network Splice*), une méthode de prédiction basée sur des réseaux neuronaux (Reese et al. 1997). Les scores varient ici de 0 à 16 ;
- MES (*MaxEntScan*), dont le calcul des scores est basé sur le principe d'entropie maximale en utilisant un large jeu de données de sites d'épissage humains (Yeo and Burge 2004). Cet outil étudie la distribution des séquences en fonction d'un ensemble de contraintes, définies par la position des nucléotides et par l'interdépendance des nucléotides adjacents et non-adjacents au sein des sites d'épissage. Ces modèles de sites d'épissage assignent un log-odd ratio (score MaxENT) sur des séquences allant de 9 (sites 5') à 23 nt (sites 3'). Ici aussi les scores varient de 0 à 16 ;
- HSF (*Human Splicing Finder*) (Desmet et al. 2009), dont les scores sont obtenus suite à l'utilisation des matrices développées par Shapiro et Senapathy en 1987 (Desmet et al. 2009). Les scores varient de 0 à 100.

En plus des sites d'épissage, les algorithmes SSF et HSF s'appliquent aux prédictions des sites de branchement. Mais d'autres algorithmes dédiés existent (Zhang 1998; Gooding et al. 2006).

L'efficacité de ces algorithmes à prédire des modifications dans la force des sites d'épissage suite à une variation a été évaluée (Houdayer et al. 2012). Les auteurs ont utilisé ces algorithmes pour évaluer, rétrospectivement, la qualité de prédiction de l'effet sur l'épissage de variations situées dans les gènes *BRCA1* et *BRCA2*, au niveau des sites consensus d'épissage décrits par Cartegni (Cartegni et al. 2002). Les scores prédictifs de ces variations ont été comparés avec les résultats expérimentaux. L'analyse des données a montré un meilleur résultat en combinant les algorithmes MES et SSF, grâce à un meilleur calcul de la sensibilité et de la spécificité pour les prédictions, par rapport à l'utilisation seule de chacun des algorithmes (Houdayer et al. 2012). Cette combinaison d'algorithmes a permis d'établir des seuils avec lesquels une sensibilité et spécificité optimales ont été obtenues. Ces seuils correspondent à la valeur de la différence entre les scores des sites consensus d'épissage au niveau de l'allèle sauvage et au niveau de l'allèle muté. Il a été proposé une différence seuil de 15% pour les scores MES et de 5% pour les scores SSF (Houdayer et al. 2012). A partir de ces seuils, des bonnes pratiques d'analyses de prédictions bio-informatiques ont été proposées (**Figure 18**). Si un variant situé au niveau ou à proximité d'un site consensus d'épissage diminue le score calculé par MES d'au moins 15% et diminue le score calculé par SSF d'au moins 5%, alors il peut être considéré comme impactant l'épissage et devrait être analysé expérimentalement en priorité. Car seule une analyse expérimentale permettra de conclure sur l'interprétation de l'impact réel du variant sur l'épissage. En revanche, pour les variants situés en dehors des régions conservées constituant les sites d'épissage, il n'a pas encore été dégagé de recommandation, ni pour les variants situés dans les autres gènes liés au syndrome HBOC.

Il est à noter que les VSI situés dans les exons 11 de ces 2 gènes ne sont pas concernés par ces recommandations. En effet, une étude portant sur une série de 108 VSI dans l'exon 11 de *BRCA1* suggère un mode de régulation spécifique de l'épissage au niveau de ces grands exons (Anczuków et al. 2008).

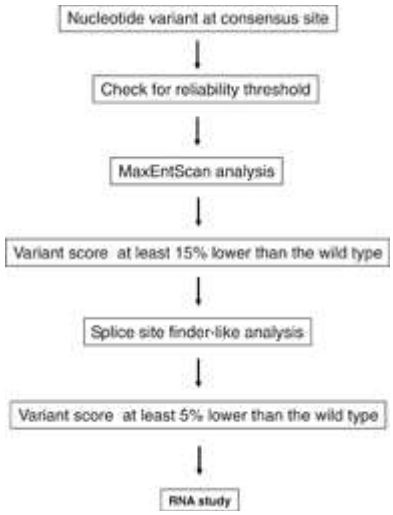

Figure 18 : Procédure pour l'analyse des scores obtenus par les algorithmes de prédictions bio-informatiques MES et SSF dédiés aux sites d'épissage (d'après Houdayer et al. 2012).

Récemment, l'équipe de Tavtigian a utilisé les scores issus de l'outil MES pour générer des probabilités à priori de pathogénicité sur l'épissage des variants des gènes *BRCA1* et *BRCA2* exoniques ou situés à proximité des sites d'épissage (-20 ; +8) (Vallée et al. 2016). Ces probabilités sont comprises entre 0,97 (variants à forte probabilité d'endommager le site donneur/accepteur ou de créer un site donneur *de novo*) et 0,02 (variants exoniques sans impact sur la jonction d'épissage et peu susceptibles de créer un site donneur *de novo*). Ces

probabilités sont utilisables pour le calcul du score de probabilité à postériori de causalité d'un VSI, intégrant l'impact potentiel d'un VSI sur l'épissage en fonction de sa position.

2. Etudes expérimentales basées sur l'analyse de l'ARN de patientes atteintes du syndrome HBOC

L'épissage de l'ARN étant en principe un mécanisme tissu-spécifique (Fortin et al. 2005), la validation d'une mutation d'épissage devrait idéalement être effectuée à partir de l'ARN issus des tissus mammaires et ovariens dans le cas des gènes *BRCA1* et *BRCA2*. Malheureusement, l'accès aux tissus des patientes n'est souvent pas possible. Aujourd'hui les analyses sont donc effectuées dans les laboratoires de diagnostic à partir des leucocytes issus du sang des patientes par RT-PCR (*reverse transcription – polymerase chain reaction*) (Houdayer et al. 2012). Cette méthode permet de comparer le profil d'épissage de la région d'intérêt, entourant la variation à tester, à celui d'individus contrôles non porteurs de variations dans le gène analysé. Les produits de RT-PCR sont séparés par migration dans un gel d'agarose et le profil d'épissage est analysé par séquençage (Gaildrat et al. 2012). Cette analyse directe des transcrits à partir de matériel biologique de patients porteurs de variations est une méthodes fiable pour établir si une variation particulière de l'ADN affecte l'épissage (Baralle et al. 2009).

Le choix des amorces est une étape importante pour la réaction de RT-PCR selon la variation à tester (Whiley et al. 2014). Le choix est simple en cas de variation exonique. En effet, au-delà de l'apparition ou du renforcement d'un épissage particulier, le séquençage des produits de PCR au niveau de la variation permet de déterminer le taux d'expression de

chacun des 2 allèles et d'en conclure si l'effet de la variation sur l'épissage est total ou partiel. Il suffit donc que les amorces se situent dans les exons entourant celui d'intérêt. Cependant cette quantification reste peu précise (Houdayer et al. 2012). En cas de variation intronique, l'analyse de cette représentation allélique n'est le plus souvent pas possible, la variation étant non détectable au niveau de l'ADNc. La recherche d'un polymorphisme à l'état hétérozygote (SNP, *single nucleotide polymorphism*) est nécessaire, afin de mesurer grâce à celui-ci un éventuel déséquilibre allélique et de conclure sur l'effet total ou partiel de la variation sur l'épissage. Si un tel SNP est situé à proximité de la variation à tester, le choix des amorces doit être fait de telle sorte à englober l'exon d'intérêt et le SNP. Si aucun SNP n'est détecté, un test fonctionnel basé sur l'utilisation de minigène est à envisager (voir plus loin) (Houdayer et al. 2012).

La méthode quantitative de pyroséquençage peut également évaluer un déséquilibre allélique, de façon plus précise (Muller et al. 2011; Gaildrat et al. 2012). Cette méthode utilise la détection de lumière produite par l'enzyme luciférase après conversion enzymatique de pyrophosphate en ATP, le pyrophosphate résultant de l'incorporation d'un dNTP sur le brin d'ADNc. Le pyroséquençage est réalisé sur les variations exoniques d'intérêt ou, de la même manière qu'en RT-PCR, englobe un SNP idéalement situé à proximité dans le cas de variations introniques. Ceci afin d'interpréter l'effet total ou partiel de la variation sur l'épissage.

Les échantillons analysés au cours du diagnostic moléculaire sont des leucocytes, mais leurs conditions de prélèvement et de conservation empêchent l'étude de l'ARN, du fait de sa fragilité. Depuis quelques années, du sang total collecté dans des tubes PAXgene®, contenant une solution permettant de stabiliser l'ARN, est utilisé pour l'étude de l'ARN (Houdayer et al.

2012). Cette méthode est très pratique pour analyser le profil d'épissage des gènes à un temps donné. Cependant, la faible expression de certains gènes dans les leucocytes rend parfois les analyses délicates. De plus, les transcrits cibles du NMD peuvent potentiellement être indétectables car aucun inhibiteur du NMD ne peut être utilisé (Houdayer et al. 2012). L'établissement de LLB à partir des leucocytes du patient, bien que plus cher, plus long et immortalisant les cellules, peut éviter ces problèmes. En effet il existe des inhibiteurs capables d'enrayer l'action du NMD et donc d'empêcher la dégradation de ces transcrits avant l'extraction de l'ARN, comme la puromycine ou la cycloheximide (Andreutti-Zaugg et al. 1997). La comparaison des profils d'expression des transcrits issus des LLB du patient, avec et sans l'utilisation de ces inhibiteurs, permet alors de conclure si l'effet de la variation sur l'épissage est total ou partiel (Houdayer et al. 2012). Mais cette approche de quantification reste peu précise.

Afin d'interpréter de façon fiable l'impact potentiel sur l'épissage d'une variation détectée chez une patiente, il est d'abord nécessaire de connaître précisément le profil physiologique d'épissage des gènes d'intérêts. Une liste d'épissages alternatifs détectés dans les gènes *BRCA1* et *BRCA2* a été récemment publiée, réalisée par des approches combinées de RT-PCR et d'électrophorèses capillaires (Colombo et al. 2014; Fackenthal et al. 2016). Récemment, de nouvelles méthodes d'analyse globale du transcriptome ont fait leur apparition, basées sur le séquençage à haut débit de l'ARNm (RNA-Seq) (Mortazavi et al. 2008). Cette technique permet potentiellement de caractériser le profil d'épissage de nombreux gènes simultanément et donc de détecter d'éventuelles anomalies d'épissage causées par des variations. La technique de RNA-Seq, en particulier de RNA-Seq ciblé, est décrite plus loin.

3. Tests fonctionnels basés sur l'utilisation de minigènes appliqués aux gènes *BRCA1* et *BRCA2*

Le test fonctionnel basé sur l'utilisation de minigènes est très utile pour l'analyse de l'effet des variations sur l'épissage. Ce test possède 3 grands avantages par rapport aux autres méthodes employées dans les laboratoires de recherche et de diagnostic moléculaire :

- Il est indépendant de la disponibilité du matériel biologique de la patiente. En effet, il ne nécessite pas d'ARN de la patiente, seulement de l'ADN génomique. Si l'ADNg du patient n'est pas disponible, la variation à tester peut être introduite dans un fragment d'ADN par mutagenèse dirigée avant insertion dans le minigène (Ho et al. 1989; Di Giacomo et al. 2013) ;
- Le test est mono-allélique, l'effet observé est donc uniquement dû à l'allèle muté. Il n'est donc pas nécessaire d'étudier un SNP en parallèle (Fredericks et al. 2015) ;
- A part le premier exon et le dernier exon des gènes étudiés, tous les exons avec les régions introniques flanquantes contenant une variation peuvent potentiellement être analysés. Les variations exoniques, les variations introniques proches et même les variations introniques profondes peuvent donc être testées (Baralle and Baralle 2005).

Comme son nom l'indique, le minigène est une version simplifiée d'un gène, cloné dans un vecteur (Baralle and Baralle 2005). Ceux utilisés pour analyser l'épissage sont classiquement constitués de 2 exons, séparés par un intron central dans lequel est introduit l'exon d'intérêt additionné des 150 premiers nucléotides introniques flanquants environ, créant alors un minigène à 3 exons. L'incorporation de l'exon central est possible par l'utilisation de sites et d'enzymes de restriction (Gaildrat et al. 2010a). Chaque plasmide ne pouvant recevoir qu'un amplicon à la fois, chaque vecteur possédera soit la région exonique

sauvage, soit la région exonique contenant la variation à tester. Ces constructions sauvages et mutantes sont ensuite transfectées en parallèle dans des cellules en culture. Les ARN des cellules transfectées sont extraits et les profils d'épissage des minigènes sauvages et mutants sont comparés par RT-PCR (Baralle and Baralle 2005; Gaildrat et al. 2010a). En revanche, les minigènes à 3 exons ne permettent pas de détecter les sauts de plusieurs exons ou les profils d'épissage complexes. Il est alors possible d'introduire 2 exons consécutifs du gène d'intérêt, en particulier lorsque l'intron entre les 2 est petit (< 300 pb). Des minigènes de ce type ont pu être étudiés pour les gènes *BRCA1* (Raponi et al. 2012) et *BRCA2* (Acedo et al. 2015).

Un grand nombre de mutations d'épissage exoniques des gènes *BRCA1* et *BRCA2*, en particulier les substitutions ponctuelles localisées dans des éléments de régulation d'épissage potentiels, a été identifié grâce à l'utilisation de ce type de minigène (Bonnet et al. 2008; Théry et al. 2011; Gaildrat et al. 2012; Di Giacomo et al. 2013; Steffensen et al. 2014). Par cette méthode, certaines variations, initialement classées comme VSI, ont pu être re-classifiées, soit comme neutres pour l'épissage, soit délétères (Bonnet et al. 2008; Gaildrat et al. 2010b; Théry et al. 2011). Notre unité de recherche a analysé avec ce test 433 VSI, exoniques et introniques, identifiés par les laboratoires du GGC, dont 174 VSI détectés chez des patients atteints du syndrome HBOC et 259 VSI détectés chez des patients atteints du syndrome de Lynch. Nous avons constaté qu'une grande fraction de VSI (29%) induit des modifications d'épissage. Nous avons eu l'occasion de comparer les résultats des tests fonctionnels basés sur l'utilisation de minigènes avec les données obtenues à partir de l'analyse de l'ARN d'un sous-groupe de patients (ARN issu du sang périphérique ou de lignées lymphoblastoïdes) et nous avons obtenu une parfaite concordance. Parmi les 433 VSI analysés, environ 18% ont eu un impact drastique sur l'épissage, comme le saut total d'un exon, et ont donc été classés comme mutations délétères. Ces mutations récemment classifiées

sont maintenant utilisées pour le diagnostic avec des conséquences importantes pour le suivi des patients, le conseil génétique et le traitement. De plus, nos données ont été intégrées dans les bases de données nationales et internationales sur les mutations des gènes du syndrome de Lynch et du syndrome HBOC (Houdayer et al. 2012; Thompson et al. 2014).

4. Interprétation de l'impact sur l'épissage des VSI

Les VSI des gènes *BRCA1* et *BRCA2* étudiés au niveau de l'ARN pour leur impact potentiel sur l'épissage sont classés selon leur effet biologique, de 1S à 3S (Houdayer et al. 2012). Les variants classés 1S n'ont pas d'effet sur l'épissage, tandis que les variants 3S ont un effet total sur l'épissage, c'est-à-dire que l'allèle muté ne produit plus qu'un transcrit alternatif ou un transcrit anormal. Les variants 2S ont un impact sur l'épissage, mais soit l'effet est partiel, soit il correspond au renforcement d'un épissage alternatif. Dans le cadre d'un effet partiel, l'allèle muté produit alors 2 transcrits : un transcrit composé des exons constitutifs du gène appelé transcrit canonique et un transcrit anormal. Dans le cadre d'un renforcement d'un épissage alternatif, l'allèle muté produit moins de transcrit canonique au profit du transcrit alternatif. Le classement définitif d'un VSI ayant un effet total sur l'épissage (classe 3S) selon les 5 classes décrites par Plon et collaborateurs (Plon et al. 2008) est alors conditionné par l'impact sur la protéine de l'épissage anormal observé :

- Un VSI synonyme ou intronique et sans effet sur l'épissage (classe 1S) sera classé non pathogène (classe 1) ;
- Un VSI faux-sens et sans effet sur l'épissage (classe 1S) restera de signification incertaine en raison de son impact potentiel sur la protéine (classe 3) ;

- Un VSI ayant un effet partiel sur l'épissage (classe 2S) restera de signification incertaine (classe 3) et des études complémentaires de co-ségrégation du variant avec la pathologie dans la famille seront nécessaires pour le classer définitivement ;
- Un VSI provoquant un saut total d'exon (classe 3S) décalant le cadre de lecture (hors phase) entrainera la production d'une protéine tronquée vraisemblablement non fonctionnelle et sera par conséquent classé délétère (classe 5). De même, un VSI provoquant le saut total d'un exon en phase au niveau d'une région codant un domaine protéique sera également classé comme délétère (classe 5).
- Un variant provoquant le saut total d'un exon (classe 3S) en phase situé en dehors d'une région codant un domaine protéique restera de signification incertaine (classe 3). A nouveau, des études complémentaires de co-ségrégation du variant avec la pathologie dans la famille seront nécessaires pour le classer définitivement

L'interprétation de l'ensemble des VSI impactant potentiellement l'épissage reste délicate, rendant parfois très difficile la classification définitive de ces variants, potentiellement utilisables pour le conseil génétique. En effet, tandis qu'il est admis que la perte d'expression totale d'un allèle peut être considérée comme délétère (Houdayer et al. 2012), il demeure une question en suspens quant au seuil à partir duquel un effet partiel peut être considéré comme délétère.

C. Objectifs des travaux de thèse

Compte-tenu de l'extrême variabilité du génome humain, chaque exome humain contenant environ 20 000 SNV (*single nucleotide variation*), dont 1000 sont rares ou très rares (fréquence < 0,1% dans la population générale) et majoritairement non répertoriées dans les bases de données, le grand défi de la génomique médicale moderne à l'ère du NGS n'est plus la détection des variations mais leur interprétation. Le défi de l'interprétation sera amplifié lors du passage progressif du séquençage de l'exome à celui du génome, chaque génome contenant environ 4 millions de SNV. Ce défi concerne l'ensemble des maladies humaines, des maladies rares, de déterminisme monogénique ou oligogénique, aux maladies communes de déterminisme complexe, associant facteurs génétiques et non génétiques.

Le problème de l'interprétation est devenu rapidement, avec l'augmentation exponentielle des tests génétiques réalisés dans le cadre de panel de gènes, la question prioritaire des laboratoires de génétique médicale assurant le diagnostic moléculaire de maladies mendéliennes. Ce problème est particulièrement aigu dans le domaine de l'oncogénétique, compte-tenu du nombre d'analyses réalisées et de l'impact médical du résultat de l'analyse génétique, et le pourcentage de VSI identifiées dans les gènes *BRCA1* et *BRCA2*, estimé environ à 30% grâce aux efforts de classification effectués depuis 2012, font de ces gènes de véritables paradigmes pour la génétique médicale moderne.

Le nombre grandissant de VSI détectés, corrélé à l'essor du séquençage à très haut débit de l'ADN, génère des données inutilisables pour le diagnostic moléculaire de la prédisposition aux cancers du sein et/ou de l'ovaire tant qu'une preuve formelle de causalité

n'est pas apportée. Les connaissances actuelles montrent qu'un grand nombre de ces VSI pourraient impacter l'épissage de l'ARN, les rendant alors délétères et utilisables dans le cadre du diagnostic. De plus, certains événements d'épissage peuvent être provoqués par des mutations introniques profondes, non recherchées en diagnostic moléculaire. L'interprétation des variations génétiques humaines et le développement de nouveaux outils permettant leur classification est un des objectifs prioritaires du Centre Normand de Génomique Médicale et de Médecine Personnalisée et dans ce cadre, la reclassification des VSI par leur étude au niveau de l'épissage dans les gènes *BRCA1* et *BRCA2*, ainsi que l'identification de mutations d'épissage dans les gènes du syndrome HBOC est l'une des priorités de notre laboratoire et de notre unité Inserm.

Le nombre de gènes connus ou suspectés d'être impliqués dans le syndrome HBOC et séquencés dans le cadre du diagnostic moléculaire d'un panel de gènes, accroit par conséquent le nombre de VSI détectés. Parmi ces VSI, la proportion de ceux impactant l'épissage dans ces nouveaux gènes n'est actuellement pas connue. L'objectif principal de ma thèse était de contribuer à évaluer cette proportion de VSI impactant potentiellement l'épissage de 9 gènes autres que *BRCA1* et *BRCA2* : *BARD1*, *BRIP1*, *CHEK2*, *PALB2*, *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2*, *XRCC3*. Nous avons émis l'hypothèse que les recommandations basées sur les prédictions bio-informatiques, émises pour les VSI situés sur les sites consensus d'épissage des exons des gènes *BRCA1* et *BRCA2* en 2012 (Houdayer et al. 2012), pouvaient être transposées et appliquées dans le cadre de ces 9 autres gènes. Ceci afin d'établir s'il est possible, à l'instar des gènes *BRCA1* et *BRCA2*, de prioriser les études des VSI d'intérêt par une approche combinatoire, mêlant prédictions bio-informatiques et tests fonctionnels d'épissage. Par conséquent nous nous sommes focalisés dans un premier temps sur l'impact potentiel sur l'épissage de VSI introniques ou exoniques synonymes situés

sur ces sites consensus. En effet, en plus d'un impact potentiel sur l'épissage, les VSI exoniques faux-sens peuvent avoir un impact combiné au niveau de la protéine du fait de la substitution d'un acide aminé par un autre. Par conséquent, afin de s'affranchir de cet impact protéique potentiel, ils n'ont pas été analysés lors de cette étude.

A partir des données de DNA-Seq, issues du diagnostic moléculaire effectué au sein du laboratoire, nous avons effectué un recueil de tous les VSI introniques et exoniques synonymes détectés au niveau des sites consensus d'épissage de nos gènes d'intérêt, hors *BRCA1* et *BRCA2*, et impactant potentiellement l'épissage de l'ARN, après une sélection basée sur les analyses de prédiction bio-informatiques. Nous avons analysé l'impact de ces VSI d'intérêt sur l'épissage par un test fonctionnel d'épissage basé sur l'utilisation de minigènes, et confirmé l'effet observé par une étude RT-PCR à partir de l'ARN des patients.

Matériel et Méthodes

Nomenclature des gènes

La numérotation des nucléotides est basée sur les séquences d'ADNc, où le c.1 représente le premier nucléotide du codon d'initiation de traduction, selon les recommandations de la Human Genome Variation Society (HGVS). Les numéros d'accession NCBI des séquences utilisées dans cette étude sont énumérés comme suit : *BARD1* (NM_000465), *BRIP1* (NM_032043), *CHEK2* (NM_001005735), *PALB2* (NM_024675), *RAD51B* (NM_133510), *RAD51C* (NM_058216), *RAD51D* (NM_002878), *XRCC2* (NM_005431), and *XRCC3* (NM_005432).

Matériel biologique

Des échantillons d'ADN génomique provenant de patients porteurs de variants impactant potentiellement l'épissage de l'ARN ont été utilisés dans le cadre des tests fonctionnels basés sur l'utilisation de minigènes (**Tableau 1**).

Tableau 1 : Liste des variants étudiées pour les tests fonctionnels basés sur l'utilisation de minigènes.

Gène	*Transcrit*	*Mutation*
BRIP1	NM_032043	c.2492+5G>A (p.?)
CHEK2	NM_001005735	c.449-5T>A (p.?)
		c.573+3A>G (p.?)
RAD51B	NM_133509	c.1036+5G>A (p.?)
RAD51C	NM_058216	c.905-3C>G (p.?)
		c.1026+5_1026+7del (p.?)

Types d'échantillons

Du sang périphérique provenant des 6 patients porteurs de variations dans les gènes *BRIP1*, *CHEK2*, *RAD51B*, *RAD51C* a été collecté pour extraction d'ADN. Cette extraction a été effectuée en utilisant l'extracteur automatique EZ1 BioRobot (Qiagen®). En parallèle, pour les patients porteurs des variants *CHEK2* c.449-5T>A et *RAD51C* c.905-3C>G, du sang périphérique a été collecté sur tube PAXgene® pour extraction d'ARN. Ces tubes contiennent un stabilisateur de l'ARN permettant de le conserver pendant 72h à température ambiante avant extraction. Cette extraction a été effectuée en utilisant l'extracteur automatique Qiacube (Qiagen®). L'ensemble des patients ont donné leur consentement éclairé à pratiquer des analyses génétiques.

Sélection des variants et prédictions bio-informatiques de l'effet de mutations sur l'épissage

Les variants introniques et exoniques synonymes, situés entre les positions -3 à +8 de la jonction exon-intron (site donneur d'épissage) et situés entre les positions -12 à +2 de la jonction intron-exon (site accepteur d'épissage) ont été sélectionnés à partir de la base de données interne du laboratoire. Seuls les variants avec une profondeur de lecture supérieure à 60x et de fréquence inférieure à 1% au laboratoire ont été conservés. Deux outils *in silico* ont été utilisés pour prédire l'effet de variants sur l'épissage de l'ARN et plus particulièrement sur la force des sites 5' et 3' d'épissage : SSF et MES. Ces algorithmes ont été interrogés simultanément par l'utilisation du logiciel Alamut (Interactive Biosoftware®). Nous avons utilisé une analyse combinée des 2 algorithmes comme décrit précédemment pour les gènes *BRCA1* et *BRCA2* (Houdayer et al. 2012), appliquée ici aux 9 gènes d'intérêt. Pour rappel, pour chaque variant, celui-ci est considéré comme probablement délétère au niveau de

l'épissage si les scores des sites consensus les plus proches, donnés par MES et SSF, diminuent d'au moins 15% et 5%, respectivement, par rapport au score sauvage. Ces outils permettent une prédiction des scores au niveau des sites consensus d'épissage ; ils ne prédisent pas le type d'épissage pouvant avoir lieu.

Tests fonctionnels basés sur l'utilisation de minigènes

Afin d'évaluer l'impact sur l'épissage des différents variants d'intérêt (**Tableau 1**), nous avons effectué des tests fonctionnels basés sur l'analyse comparative du profil d'épissage de type sauvage et mutant au niveau de l'exon d'intérêt par l'utilisation de minigènes. Ces minigènes ont été préparés en utilisant le vecteur pCAS2. Ce vecteur porte deux exons (appelés ici A et B) avec une séquence dérivée du gène *SERPING1*/*C1NH* humain, séparés par un intron contenant des sites de clonage BamHI et MluI. L'expression du minigène pCAS2 est sous le contrôle d'un promoteur CMV.

Préparation des séquences comprenant les variants à tester

Chaque région d'intérêt a été amplifiée par PCR à partir de l'ADN génomique de la patiente. Les amorces sont spécifiques de la séquence à tester, de telle sorte à ce que l'exon d'intérêt et les 150 premiers nucléotides des introns entourant l'exon soient amplifiés (voir **Annexe I**). Sur chaque amorce sens, en 5', a été ajoutée la séquence GACCGGATCC, la séquence soulignée correspondant au site de restriction de l'enzyme BamHI. Sur chaque amorce antisens, en 5', a été ajoutée la séquence GACCACGCGT, la séquence soulignée correspondant au site de restriction de l'enzyme MluI. Pour les 2 variants situés dans le gène *CHEK2*, l'intron 4 faisant seulement 118 pb, la séquence d'intérêt à amplifier contenait la

région allant des 150 derniers nucléotides de l'intron 3 jusqu'au 150 premiers nucléotides de l'intron 5, pour englober les exons 4 et 5.

Chaque variant a été analysé séparément. La PCR a été effectuée par l'utilisation du kit AmpliTaq Gold® DNA Polymerase with Buffer II and $MgCl_2$ (Applied Biosystems®). Le mélange réactionnel comprenait 2 µl du tampon Buffer II (10X), 1.6 µl de chlorure de magnésium $MgCl_2$ à 25 mM, 1.6 µl de nucléotides dNTP GeneAmp® dNTP (100 mM) Blend (Applied Biosystems®) préalablement dilués à 10 mM, 0.2 µl de chaque amorce sens et antisens à 50 µM, 0.17 µl d'AmpliTaq Gold® DNA Polymerase (5 U/µl), 100 ng d'ADN génomique et de l'eau stérile, pour un volume total de 18 µl. L'activation de l'enzyme s'effectue à 95°C pendant 10 min. L'amplification s'effectue pendant 26 cycles comprenant une phase de 20 sec à 94°C, une phase de 20 sec à 63°C (avec une diminution de 0,5°C par cycle jusqu'à 50°C), une phase de 45 sec à 72°C, puis pendant 20 cycles supplémentaires comprenant une phase de 20 sec à 94°C, une phase de 20 sec à 50°C et une phase de 45 sec à 72°C. L'élongation finale des brins s'effectue pendant 7 min à 72°C. La température est finalement abaissée et maintenue à 4°C. La taille attendue des produits de RT-PCR variait entre 445 et 1 183 pb. Ils ont été visualisés en faisant migrer 4 µl de mélange réactionnel par électrophorèse sur gel d'agarose à 1,5% contenant l'intercalant de l'ADN SYBR Safe (Life Technologies®) et par l'utilisation du système d'acquisition d'images Gel Doc EZ System (Bio-Rad®). La migration a été effectuée en présence d'un marqueur de taille 100 pb DNA Molecular weight Marker XIV (Roche®) pendant 45 min sous une tension de 100 volts. Chaque bande a été prélevée, purifiée par l'utilisation du kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel®), éluée dans 30µl du tampon d'élution du kit avant d'être séquencée par l'utilisation du kit de séquençage BigDye® Terminator v1.1 Cycle Sequencing Kit sur séquenceur automatique ABI 3130xl (Applied Biosystems®). Le mélange réactionnel

comprenait 2.4 µl du tampon Sequencing Buffer (5X), 0.6 µl de BigDye® Terminator v1.1 Ready Reaction Mix, 6 µl d'amorce sens ou antisens à 0,5 µM et 1 µl d'amplicons purifiés.

Après s'être assuré de la conformité des séquences et de la détection de chaque variant, les amplicons (inserts) ont subi une double digestion par les enzymes de restriction BamHI et MluI, dans un volume réactionnel de 50 µl incubé à 37°C pendant 2h et contenant 5 µl du tampon NEBuffer 3.1 (10X) (New England Biolabs®), 0.5 µl d'enzyme BamHI (20 U/µl) (New England Biolabs®), 1 µl d'enzyme MluI (10 U/µl) (New England Biolabs®), 28 µl du produit de PCR purifié et 15.5 µl d'eau stérile. Les produits de digestion ont ensuite été purifiés par l'utilisation du kit NucleoSpin Extract II (Macherey-Nagel®). Après les étapes de lavage, les colonnes ont été séchées avec le couvercle ouvert pendant 5 min à 70°C et les produits ont été élués après incubation pendant 3 min à température ambiante dans 25 µl de tampon d'élution. Ils ont enfin été dosés par analyse spectrophotométrique (Nanovue®).

Préparation du minigène pCAS2

Le plasmide (vecteur) pCAS2 a subi une double digestion par les enzymes de restriction BamHI et MluI, dans un volume réactionnel de 100 µl incubé à 37°C pendant 2h et contenant 10 µl du tampon NEBuffer 3.1 (10X) (New England Biolabs®), 3 µl d'enzyme BamHI (20 U/µl) (New England Biolabs®), 6 µl d'enzyme MluI (10 U/µl) (New England Biolabs®), 10 µg du plasmide pCAS2 et de l'eau stérile. Les produits de digestion ont ensuite été incubés pendant 15 min à 65°C afin d'inactiver les enzymes de restriction. Le vecteur digéré a été déphosphorylé par l'utilisation du kit Shrimp Alkaline Phosphatase (SAP) (Affimetrix®) pendant 10 min à 37°C puis pendant 15 min à 65°C. Le vecteur a été visualisé en faisant migrer la totalité du mélange réactionnel dans 4 puits par électrophorèse sur gel

d'agarose à 1% contenant du bromure d'éthidium pendant 45 min sous une tension de 100 volts et par l'utilisation du système d'acquisition d'images Gel Doc XR (Bio-Rad®). Le vecteur a ensuite été purifié par l'utilisation du kit NucleoSpin Extract II (Macherey-Nagel®) selon les recommandations du fournisseur, puis dosé par analyse spectrophotométrique (Nanovue®).

Clonage des inserts au sein du vecteur minigène pCAS2

Ligation

Quinze à trente-cinq nanogrammes de chaque produit de PCR digéré et purifié ont été insérés par ligation dans 80 ng de vecteur minigène pCAS2 par l'utilisation du kit T4 DNA Ligase (New England Biolabs®) dans 10 µl de mélange réactionnel. Ce mélange a été incubé pendant 1h à température ambiante. Un vecteur supplémentaire n'a été mis en contact avec aucun produit de PCR afin de l'utiliser comme contrôle négatif tout au long du protocole.

Transformation

Le mélange de ligation a été ajouté à 100 µl d'une suspension de bactéries compétentes XL1 Blue. Ce mélange a été incubé sur la glace pendant 20 min. Un choc thermique a été appliqué aux bactéries (42°C pendant 45 sec puis 2 min sur la glace) afin de permettre l'intégration des plasmides par les bactéries. Un millilitre de milieu LB (*lysogeny broth*) a été ajouté au mélange et le tout a été incubé 1h à 37°C sous agitation. Ce mélange a ensuite été centrifugé pour éliminer environ 1 ml de surnageant et remettre les bactéries en suspension dans les 100 µl restants, avant d'ensemencer des boîtes de Pétri (1 boîte par insert, mélange d'amplicons sauvages et mutants) coulées à l'agar-LB et contenant 50 µg/ml de carbénicilline. Les bactéries ont été cultivées pendant une nuit à 37°C.

Sélection des plasmides

Dans chaque boîte de Pétri, 6 colonies distinctes de bactéries ont été sélectionnées et chacune a été remise en culture dans un tube contenant 3 ml de milieu de culture LB supplémenté de 50 µg/ml de carbénicilline, pendant une nuit à 37°C. Les tubes ont été centrifugés et le surnageant a été éliminé. L'ADN plasmidique a ensuite été extrait par l'utilisation du kit MiniPrep NucleoSpin® Plasmid (Macherey-Nagel®) selon les recommandations du fournisseur. L'ADN plasmidique issu des 6 colonies bactériennes pour chaque variant testé a été séquencé par l'utilisation du kit de séquençage BigDye® Terminator v1.1 Cycle Sequencing Kit sur séquenceur automatique ABI 3130xl, (Applied Biosystems®). Le mélange réactionnel comprenait 1.5 µl du tampon Sequencing Buffer (5X), 1 µl de BigDye® Terminator v1.1 Ready Reaction Mix, 1 µl d'amorce Seq-Ins-F à 3.2 µM (voir **Annexe II**), 1 µl d'ADN plasmidique et 6.5 µl d'eau stérile. Le séquençage a permis de vérifier le bon clonage des inserts et de sélectionner, pour chaque variant testé, un clone comportant la séquence sauvage et un clone comportant la séquence mutée. Pour chaque variant testé, ces 2 clones ont été dosés par analyse spectrophotométrique (Nanovue®) et normalisés chacun pour obtenir 400 ng d'insert dans 5 µl.

Préparation des cellules pour la transfection

Le contenu d'une flasque de cellules HeLa à confluence a été vidé pour ne garder que les cellules. Celles-ci ont été lavées par 10 ml de tampon phosphate salin 1X stérile (PBS, *Phosphate Buffered Saline*) en prenant garde à ne pas les décrocher de la paroi de la flasque. Les cellules ont ensuite été décrochées par 500µl de trypsine recouvrant bien toute la paroi pendant 5 min à 37°C. Un volume de 9.5 ml de milieu RPMI (*Roswell Park Memorial Institute medium*) contenant 10% de sérum de veau fœtal (SVF) a été ajouté dans la flasque et

mélangé en croix. Le liquide a été transféré dans un tube Falcon de 50 ml et les cellules ont été dosées et diluées jusqu'à obtenir une suspension de cellules à la concentration de 8.10^4 cellules/ml. Un millilitre de cette suspension a été transféré dans chaque puits de plaques contenant 12 puits pour transfection. Le tout a été incubé à 37°C jusqu'à obtenir une confluence en cellules d'environ 60%.

Transfection des minigènes au sein des cellules HeLa

Pour chaque variant testé, les minigènes sauvages et mutants (0,4 µg/puits) ont été transfectés en parallèle dans les cellules HeLa arrivées à environ 60% de confluence dans les plaques 12 puits en utilisant le kit FuGENE 6 Transfection Reagent (Roche®), pendant 24h à 37°C.

Collecte des cellules, extraction de l'ARN total et RT-PCR

Vingt-quatre heures après la transfection, Le milieu de culture de chaque puits a été retiré et les cellules ont été lavées avec du PBS 1X. L'ARN total a été extrait des cellules en utilisant le kit NucleoSpin RNA II (Macherey Nagel®) selon les recommandations du fournisseur avant d'être dosé par analyse spectrophotométrique (Nanovue®). Les transcripts des minigènes ont été analysés par RT-PCR semi-quantitative dans un volume réactionnel de 25 µl en utilisant le kit OneStep® RT-PCR (Qiagen®), 200 ng d'ARN total et les amorces sens (pCAS-KO1-F) et antisens (pCAS-2-R) spécifiques du minigène pCAS2, à 3.2 µM (voir **Annexe II**). La rétrotranscription est réalisée pendant 30 min à 50°C. La reverse-transcriptase est inactivée pendant 15 min à 95°C tandis que la polymérase est activée. L'amplification s'effectue pendant 30 cycles comprenant une phase de 30 sec à 94°C, une phase de 30 sec à 57°C, une phase de 1 min à 72°C. L'élongation finale des brins s'effectue pendant 10 min à

72°C. La température est finalement abaissée et maintenue à 4°C. Les 25 µl de produits de RT-PCR ont subi une migration par électrophorèse sur gel d'agarose à 2,5% contenant du bromure d'éthidium pendant 40 min sous une tension de 100 volts et visualisés par exposition à la lumière ultraviolette dans des conditions saturantes et non saturantes en utilisant le système d'acquisition d'images Gel Doc XR (Bio-Rad®). Les bandes ont ensuite été purifiées par l'utilisation du kit NucleoSpin Extract II (Macherey-Nagel®) selon les recommandations du fournisseur. Le séquençage a été effectué par l'utilisation du kit BigDye® Terminator v1.1 Cycle Sequencing sur séquenceur automatique ABI 3130xl, (Applied Biosystems®). Le mélange réactionnel comprenait 1.5 µl du tampon Sequencing Buffer (5X), 1 µl de BigDye® Terminator v1.1 Ready Reaction Mix, 1 µl d'amorce pCAS-Seq-F ou pCAS-Seq-R à 3.2 µM (voir **Annexe II**), 3 µl de produits de PCR et de l'eau stérile, pour un volume total de 10 µl.

Analyses des transcrits par RT-PCR

Pour les patients porteurs des variants *CHEK2* c.449-5T>A et *RAD51C* c.905-3C>G, les transcripts d'intérêt ont été analysés par RT-PCR semi-quantitative dans un volume réactionnel de 50 µl en utilisant le kit OneStep® RT-PCR (Qiagen®), 200 ng d'ARN total et les amorces sens et antisens situées respectivement dans l'exon 3 et l'exon 6 du gène *CHEK2*, ainsi que dans l'exon 6 et l'exon 8 du gène *RAD51C*, à 0,8 µM (voir **Annexe III**). La rétrotranscription est réalisée pendant 30 min à 50°C. La reverse-transcriptase est inactivée pendant 15 min à 95°C tandis que la polymérase est activée. L'amplification s'effectue pendant 26 cycles comprenant une phase de 20 sec à 94°C, une phase de 20 sec à 63°C (avec une diminution de 0,5°C par cycle jusqu'à 50°C), une phase de 45 sec à 72°C, puis pendant 20 cycles supplémentaires comprenant une phase de 20 sec à 94°C, une phase de 20 sec à

50°C et une phase de 45 sec à 72°C. L'élongation finale des brins s'effectue pendant 7 min à 72°C. La température est finalement abaissée et maintenue à 4°C. La taille attendue des produits de RT-PCR variait entre 153 et 372 pb. Ils ont été visualisés en faisant migrer 4 µl de mélange réactionnel par électrophorèse sur gel d'agarose à 1,5% contenant l'intercalant de l'ADN SYBR Safe (Life Technologies®) et par l'utilisation du système d'acquisition d'images Gel Doc EZ System (Bio-Rad®). La migration a été effectuée en présence d'un marqueur de taille 100 pb DNA Molecular weight Marker XIV (Roche®) pendant 45 min sous une tension de 100 volts. Chaque bande a été prélevée, purifiée par l'utilisation du kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel®), éluée dans 30µl du tampon d'élution du kit avant d'être séquencée par l'utilisation du kit de séquençage BigDye® Terminator v1.1 Cycle Sequencing Kit sur séquenceur automatique ABI 3130xl (Applied Biosystems®). Le mélange réactionnel comprenait 2.4 µl du tampon Sequencing Buffer (5X), 0.6 µl de BigDye® Terminator v1.1 Ready Reaction Mix, 6 µl d'amorce sens ou antisens à 0,5 µM et 1 µl d'amplicons purifiés.

Résultats

Dans cette étude, nous avons analysé la base de données interne de variants du laboratoire, à partir d'une cohorte de 3 181 patients précédemment séquencés dans le cadre du diagnostic moléculaire du syndrome HBOC, sur un panel de gènes (Castéra et al. 2014). Nous avons recueilli l'ensemble des variants introniques et exoniques synonymes rares situés sur les sites consensus d'épissage de 9 gènes liés au syndrome HBOC et autres que *BRCA1* et *BRCA2*: *BARD1*, *BRIP1*, *CHEK2*, *PALB2*, *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2*, et *XRCC3*. Nous avons ensuite émis l'hypothèse que les recommandations émises en 2012 (Houdayer et al. 2012) et basées sur les prédictions bio-informatiques pour les gènes *BRCA1* et *BRCA2* pouvaient s'appliquer à ces autres gènes. Les variants faisant diminuer les scores de plus de 15% pour MES et de 5% pour SSF ont été sélectionnés de par leur effet potentiel sur l'épissage. Nous avons recueilli 35 variants, dont 17 uniques, situés sur les sites 5' et 3' d'épissage des gènes d'intérêt (**Tableau 2**). Parmi ces variants, 10 étaient considérés comme impactant directement l'épissage car ils correspondaient à des destructions de sites canoniques donneur ou accepteur d'épissage (AG/GT) (Houdayer et al. 2012). Les 7 variants restants sont listés ici : *BRIP1* c.2492+5G>A, *CHEK2* c.449-5T>A, *CHEK2* c.573+3A>G, *RAD51B* c.452+3A>G, *RAD51B* c.1036+5G>A, *RAD51C* c.905-3C>G et *RAD51C* c.1026+5_1026+7del.

Gène/ Transcrit	Mutation	Exon /Intron	Fréquence chez les contrôles (données ExAC)	Score sauvage > Score muté (% variation)		Histoire personnelle (âge au diagnostic)	Histoire familiale (âge au diagnostic)
				Max-EntScan	Splice Site Finder-like		
BARD1 NM_000465	c.159-1G>T	1	-	7.47 > 0 (100%)	81.54 > 0 (100%)	CS bilatéral (54, 56)	Neveu, TC (31) ; Tante maternelle, CO (45) ; Tante maternelle, CP (78)
	c.2001+1G>C	10	1.5e-5	3.50 > 0 (100%)	68.92 > 0 (100%)	CS (39)	Tante maternelle[§], CS (55)
BRIP1 NM_032043	c.1474-2A>G	10	-	5.19 > 0 (100%)	84.90 > 0 (100%)	CS (55)	Cousine paternelle, CS (<50) ; Cousine paternelle, CS (63) ; Cousine paternelle, CE (65)
CHEK2 NM_001005735	c.449-2A>G	3	-	7.66 > 0 (100%)	92.89 > 0 (100%)	I: Homme, CS (52) II: CS (35)	I: - II: Père, CF (65)
	c.573+1G>A	4	1.7e-4	8.10 > 0 (100%)	80.65 > 0 (100%)	CS (40)	-
	c.1590+2T>C	14	-	8.65 > 0 (100%)	85.56 > 78.76 (7.9%)	CS (48)	Tante paternelle, CS (40)
PALB2 NM_024675	c.2748+2T>G	7	-	11.08 > 0 (100%)	95.95 > 0 (100%)	CS (38 et 49)	Mère, TC (36) ; Grand-père maternel, CP
RAD51C NM_058216	c.705+1G>T	4	-	9.11 > 0 (100%)	84.19 > 0 (100%)	CS (37)	Mère, CS (57) ; Grande tante maternelle, CU (53)
	c.706-2A>G	4	3e-5	11.10 > 0 (100%)	86.72 > 0 (100%)	I: CS bilatéral (31, 50) II: CO (66) III: CO (50) IV: CS (41) V: CO (51)	I: Sœur, CO (56) ; Tante maternelle, CU (51) ; Grand-mère maternelle, CS (38) II: Sœur, CO (70) ; Mère, CO (76) ; Tante maternelle, CO (85) III: - IV: 2 sœurs, CS (33, 42) ; 5 tantes maternelles, CS (?, ?, 58, 59, 79) ; 2 cousines maternelles, CS (?, 43) ; Tante paternelle, CS (68) V: Tante maternelle, CS (60)
	c.706-2A>T	4	-	11.10 > 0 (100%)	86.72 > 0 (100%)	CS (32)	Tante maternelle, CS (51)

BRIP1 NM_032043	c.2492+5G>A	17	1.6e-5	9.66 > 6.62 (31.5%)	95.87 > 83.72 (12.7%)	CS (34)	Grand-père paternel, CP (70) ; Grande tante paternelle, CS (50)
CHEK2 NM_001005735	c.449-5T>A	3	8e-4	7.66 > 5.95 (22.3%)	92.89 > 86.48 (6.9%)	I: CS (48) II: CS (63) III: CS (47) IV: Homme, CP (48) V: CS (41) VI: CS (31) VII: CS (56) VIII: CO (66) IX: CS (46) X: non atteinte XI: CS (38) XII: CS (41, 67)	I: - II: 2 sœurs, CS (54, 70) ; Mère, CS (77) III: Mère, CS (40) ; CS (63) IV: Sœur, CS (<40) ; Mère, CS (45) V: 2 sœurs, CS (36, 58) VI: Tante paternelle, TC (49) VII: Mère, BC (40) ; Tante maternelle, CS (56) ; Grand-mère paternelle, CU VIII: - IX: - X: Mère, CO (46) ; Grand-mère maternelle, CO (60) ; Tante paternelle, CS (68) XI: LLC (55) XII: -
	c.573+3A>G	4	-	8.10 > 5.37 (33.7%)	80.65 > 76.29 (5.4%)	CS bilatéral (39, 44)	-
RAD51B NM_133509	c.452+3A>G	5	3e-5	9.11 > 4.59 (49.6%)	84.19 > 79.83 (5.2%)	CS bilatéral (45, 57)	Grand-mère, CS (<50)
	c.1036+5G>A	10	7e-4	10.57 > 8.38 (20.7%)	94.36 > 82.21 (12.9%)	I: CS (44) II: CS (31) III: CO (54)	I: Sœur, CS (65) ; 2 tantes maternelles, CS II: Oncle paternel, CE (50) ; Grand-père paternel, CE (44) ; 3 grandes tantes paternelles, CS (50, 55, 68) III: Tante maternelle, CS (67)
RAD51C NM_058216	c.905-3C>G	6	-	8.18 > 0 (100%)	97.08 > 86.26 (11.1%)	CO (56)	Grand-mère paternelle, CS (50)
	c.1026+5 _1026+7del	8	-	1.98 > 0 (100%)	77.67 > 65.28 (16%)	CS (59)	Mère, CE (72)

Tableau 2 : Caractéristiques personnelles et familiales des patients porteurs de variants introniques impactant potentiellement l'épissage de l'ARN par les prédictions bio-informatiques (MES, SSF). Les variants localisés sur les sites canoniques AG/GT sont représentés en gris foncé. Les autres VSI localisés au niveau des sites 5' et 3' d'épissage sont représentés en blanc. ExAC: Exome Aggregation Consortium ; CS: cancer du sein ; TC : tumeur cérébrale ; CP: cancer du pancréas ; CE: cancer de l'estomac ; CF : cancer du foie ; CU: cancer de l'utérus ; CO: cancer de l'ovaire ; CLL: leucémie lymphoïde chronique. § Cette tante maternelle est aussi porteuse du variant *BARD1* c.2001+1G>C.

Dans le but d'évaluer l'impact de ces variants sur l'épissage, nous avons effectué pour chacun des tests fonctionnels basés sur l'utilisation de minigènes (Soukarieh et al. 2016). Nous avons amplifié chaque région génomique d'intérêt (exon et au moins 150 pb d'introns juxta-exoniques) à partir de l'ADN génomique des patients et cloné chaque région dans le plasmide rapporteur pCAS2 afin de générer des minigènes hybrides à 3 exons, sauvages (wt, *wild-type*) et mutants (**Figure 19**). Pour les 2 variants situés dans le gène *CHEK2*, l'intron 4 ne faisant que 118 pb, la séquence d'intérêt comprenait en conséquence les exons 4 et 5, afin de générer un minigène hybride à 4 exons. Les plasmides ont été transfectés dans des cellules HeLa, l'ARN a été extrait 24h après transfection et le profil d'épissage des minigènes a été observé après RT-PCR et séquençage des produits obtenus (Gaildrat et al. 2010a).

Figure 19 : Représentation schématique des constructions minigènes pCAS2 utilisé lors du test fonctionnel. Les exons d'intérêt (boîtes noires) et les régions juxta-exoniques (lignes noires) ont été clonées dans le minigène pCAS2, en utilisant les sites de restrictions enzymatiques BamHI et MluI, comme décrit précédemment (Soukarieh et al. 2016). Les flèches épaisses représentent le promoteur du cytomégalovirus. La localisation des amorces (pCas-KO1F et pCAS-2-R) utilisées pour les analyses par RT-PCR et séquençage sont représentées par les petites flèches (voir **Annexe II**).

Nos résultats ont montré que sur les 7 variants testés ici, 2 causaient un saut total de l'exon (**Figure 20**). En effet, le variant *RAD51C* c.1026+5_1026+7del cause la production d'un transcrit aberrant correspondant au saut de l'exon 8 (r.966_1026del). Cette altération conduit à l'introduction d'un codon STOP prématuré et donc à la production d'une protéine instable ou tronquée et vraisemblablement non fonctionnelle (p.Arg322Serfs*22).

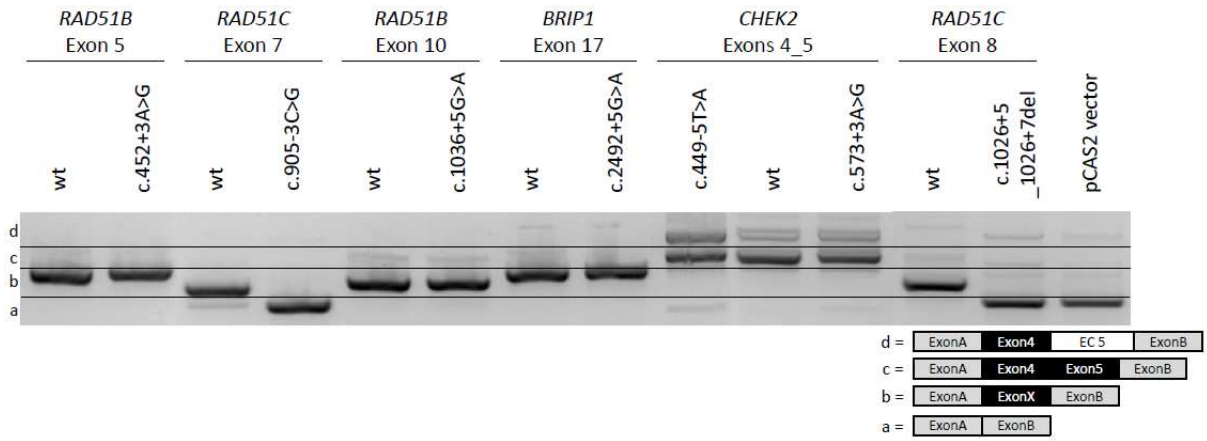

Figure 20 : Analyse des profils d'épissage des minigènes sauvages et mutants. Les transcrits ont été analysés par RT-PCR et séquençage, après expression dans des cellules HeLa et en utilisant les amorces pCas-KO1F et pCAS-2-R (voir **Annexe II**). Les produits de RT-PCR ont été séparés par électrophorèse sur gel d'agarose contenant du bromure d'ethidium et par le séquençage des différentes bandes prélevées et purifiées. Chaque épissage alternatif est illustré en bas à droite et correspond à une zone du gel (a à d), délimitée par des lignes noires horizontales. EC : exon cryptique.

Le second variant ayant un impact majeur sur l'épissage était le variant c.905-3C>G, situé dans l'intron 6 du gène *RAD51C*. Ce variant cause la production d'un transcrit alternatif correspondant au saut de l'exon 7 (r.905_965del) (**Figure 20**). Cet épissage alternatif conduit à l'introduction d'un codon STOP prématuré et donc à la production d'une protéine instable ou tronquée et vraisemblablement non fonctionnelle (p.Glu303Trpfs*41). Ce transcrit alternatif est aussi faiblement produit par la construction sauvage. Nous avons confirmé ces observations par RT-PCR et séquençage à partir de l'ARN du patient porteur de ce variant et de 2 contrôles (**Figure 21**). En effet, l'analyse a montré que le transcrit correspondant au saut de l'exon 7 est détecté chez les contrôles (données non montrées), indiquant donc un épissage alternatif. Chez le patient, nous avons détecté 2 transcrits, correspondant au transcrit sauvage et au transcrit correspondant au saut de l'exon 7. Ces 2 transcrits étaient quantitativement

comparables, démontrant un défaut majeur d'épissage conformément aux recommandations (Houdayer et al. 2012).

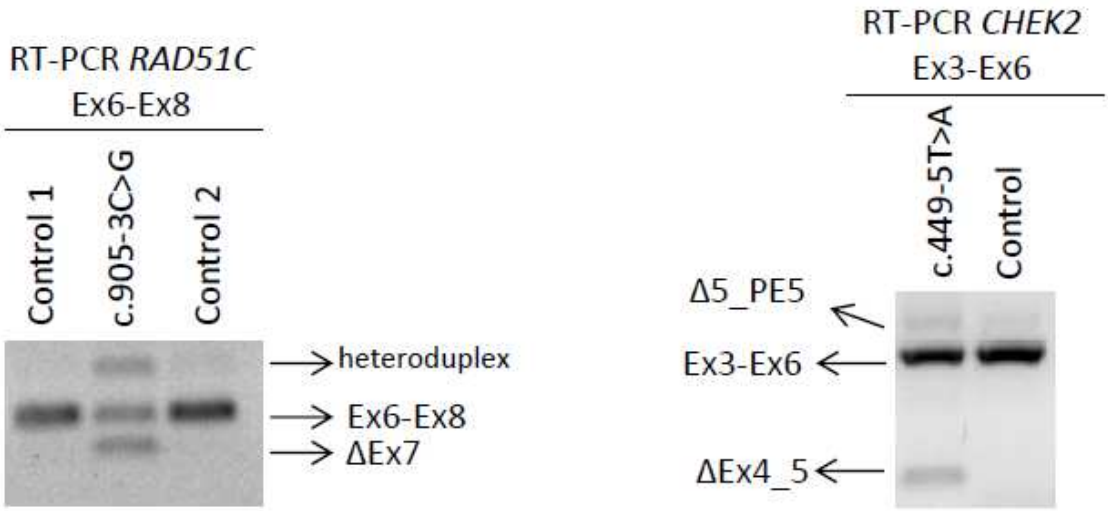

Figure 21 : Analyse par RT-PCR des transcrits des gènes *RAD51C* et *CHEK2* à partir de l'ARN de patients, issu de sang périphérique collecté sur tubes PAXgene®. Le patient porteur du variant *RAD51C* c.905-3C>G a été testé avec 2 autres contrôles négatifs. Ex6-Ex8 : transcrits comprenant les exons 6 à 8 du gène *RAD51C*. ΔEx7: transcrit correspondant au saut de l'exon 7. Le patient porteur du variant *CHEK2* c.449-5T>A a été testé avec un autre contrôle négatif. Ex3-Ex6 : transcrits comprenant les exons 3 à 6 du gène *CHEK2*. ΔEx4_5 : transcrit correspondant au saut combiné des exons 4 et 5. Δ5_PE5: transcrit correspondant au saut de l'exon 5 combiné à l'inclusion d'un exon cryptique dans l'intron 5.

Les variants *BRIP1* c.2492+5G>A, *RAD51B* c.452+3A>G et *RAD51B* c.1036+5G>A n'ont pas d'effet visible sur l'épissage (**Figure 20**). Les variants c.449-5T>A et c.573+3A>G, situés dans l'intron 3 et dans l'intron 4 du gène *CHEK2*, respectivement, ont un effet partiel sur l'épissage (**Figure 20**). En effet, ces 2 variants causent la production faible d'un transcrit supplémentaire aberrant correspondant au saut combiné des exons 4 et 5 (r.449_721del). Cette

altération produit une délétion en phase (p.Lys150_Val241delinsIle). Les 2 autres transcrits sont produits à la fois par le minigène sauvage et le minigène mutant. Le transcrit majoritaire correspondait au transcrit sauvage. Le transcrit minoritaire correspondait à un épissage alternatif combinant le saut de l'exon 5 et l'inclusion d'un exon cryptique de 359 nucléotides localisé au début de l'intron 5 (r.574_721del; 721_722ins721+54_721+412) (données non montrées). Nous avons confirmé ces observations par RT-PCR et séquençage à partir de l'ARN du patient porteur du variant c.449-5T>A et d'un contrôle négatif (**Figure 21**). En effet, l'analyse a montré que le transcrit correspondant au saut de l'exon 5 combiné à l'inclusion de l'exon cryptique est détecté chez le contrôle (données non montrées), indiquant donc un épissage alternatif. Chez le patient, nous avons détecté un transcrit supplémentaire, d'intensité plus faible que le transcrit sauvage, correspondant au saut combiné des exons 4 et 5, et suggérant un effet partiel sur l'épissage.

Ces travaux sont en préparation en vue d'une soumission sous forme de *Short Communication* dans ***Human Mutation***.

Splicing alterations in non *BRCA* genes contribute to HBOC syndrome

Grégoire Davy[1,2], Hélène Tubeuf[2,3], Laurent Castéra[1,2], Antoine Rousselin[1,2], Nicolas Goardon[1,2], Pascaline Gaildrat[2], Thierry Frebourg[2,4], Dominique Vaur[1,2], Alexandra Martins[2†], Sophie Krieger[1,2,5*†]

[1]Department of Cancer Biology and Genetics, CLCC François Baclesse, Normandy Centre for Genomic and Personalized Medicine, Caen, France; [2]INSERM U1079, Institute for Research and Innovation in Biomedicine, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; [3]Interactive Biosoftware, Rouen, France; [4]Department of Genetics, University Hospital, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; [5]University of Caen, France.

[*] To whom correspondence should be addressed at: Laboratoire de Biologie et Génétique du Cancer, Centre de Lutte Contre le Cancer François Baclesse, 3 avenue du Général Harris, 14076 Caen cedex, France. Tel: +33 231455195; Email: s.krieger@baclesse.unicancer.fr

† Co-last authors.

Keywords: Hereditary breast and ovarian cancer, minigene assays, RNA splicing mutation, bioinformatics predictions

Discussion et Perspectives

Le nombre d'études sur l'épissage de l'impact de VSI détectés dans les gènes du syndrome HBOC autres que *BRCA1* et *BRCA2* est encore limité. Dans l'optique d'évaluer le taux de mutations d'épissage introniques situées au niveau des sites consensus d'épissage de 9 gènes liés au syndrome HBOC (*BARD1*, *BRIP1*, *CHEK2*, *PALB2*, *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2*, *XRCC3*), nous avons analysé les données de séquençage à haut débit de l'ADN de 3 181 patients, testés au laboratoire pour le diagnostic moléculaire du syndrome HBOC. Nous nous sommes focalisés sur les VSI introniques et synonymes situés au niveau des régions conservées. Nous avons utilisé une approche de sélection bio-informatique des variants impactant potentiellement l'épissage, couplée à des tests fonctionnels basés sur l'utilisation de minigènes. L'ensemble des résultats a souligné l'apport de cette stratégie combinée dans le cadre de l'interprétation de l'effet des variations génétiques sur l'épissage. En effet, 57% des variants testés, hors sites canoniques AG/GT, avaient un effet sur l'épissage.

Nous avons d'abord déterminé que 35 patients (1,1%) étaient porteurs d'une variation intronique rare localisée au niveau des sites consensus donneurs et accepteurs d'épissage dans nos neuf gènes d'intérêt. Il n'a pas été observé de VSI synonyme impactant potentiellement l'épissage. Parmi ces patients, 14 patients (40%) étaient porteurs de mutations touchant les sites canoniques d'épissage AG/GT. Ces mutations, abolissant ou abaissant drastiquement la force des sites d'épissage, sont considérés d'emblée comme ayant un effet total sur l'épissage. Cependant, il convient de s'assurer qu'il n'existe pas de site cryptique à proximité du site susceptible de conduire à la production d'un transcrit en phase (Houdayer et al. 2012) avant de les interpréter comme potentiellement délétères. En dehors de la mutation c.706-2A>G dans le gène *RAD51C*, déjà décrite comme une mutation délétère d'épissage (Golmard et al. 2013), toutes les autres mutations sont décrites ici pour la première fois. De plus, d'après la

base de données en ligne Exome Agrégation Consortium (ExAC) sur la population non-finlandaise européenne (Lek et al. 2016), ces mutations sont extrêmement rares (**Tableau 2**). Il est à noter que la mutation c.706-2A>G est l'une des plus fréquentes dans notre série (5 patients soit 0,16%). Cette mutation a été détectée dans 2 familles ne comportant que des cancers de l'ovaire et dans une famille avec cancers du sein et de l'ovaire, suggérant que l'inactivation de la fonction de RAD51C peut entraîner les deux types de cancers (Meindl et al. 2010; Neidhardt et al. 2016). Fait intéressant, cette mutation a également été détectée dans une famille ne comportant que des cancers du sein, ce qui ne correspond pas au phénotype clinique décrit actuellement. En effet il existe très peu de cas de mutations de *RAD51C* associées uniquement à des cancers du sein (Osorio et al. 2012; Schnurbein et al. 2013). Des études de co-ségrégation semblent nécessaires afin de déterminer l'implication de ce gène dans des syndromes de cancers du sein isolés. A ce jour, les mutations de *RAD51C* ne confèreraient pas un risque augmenté de cancer du sein (Sopik et al. 2015). En effet, des études complémentaires vont être réalisées dans le cadre d'une étude portée par le GGC (étude Tumospec). Cette étude visera à préciser les phénotypes associés dans les familles porteuses de mutations délétères dans les gènes liés au syndrome HBOC ainsi que de préciser les risques, par des études de co-ségrégation. Ceci permettra d'apporter des arguments pour émettre des recommandations cliniques, en termes de surveillance et de chirurgie prophylactique éventuelle. Pour le gène *RAD51C*, il semble se dégager de la littérature que les porteuses de mutations ont un risque de cancer de l'ovaire suffisant pour justifier la proposition d'une ovariectomie après 50 ans (Sopik et al. 2015). A ce jour, cette proposition est étudiée au sein du GGC. De plus, il n'y a pas de recommandations concernant le cancer du sein.

Dans cette étude, nous avons analysé 7 variants, sélectionnés d'après les prédictions bio-informatiques pour leur impact potentiel sur l'épissage de 4 gènes liés au syndrome HBOC autres que les gènes *BRCA1* et *BRCA2* : *BRIP1*, *CHEK2*, *RAD51B*, *RAD51C*. Cette analyse a été effectuée par test fonctionnel basé sur l'utilisation de minigènes. Parmi ces variants testés, 3 n'avaient pas d'effet sur l'épissage (c.2492+5G>A, *RAD51B* c.452+3A>G et *RAD51B* c.1036+5G>A), malgré des prédictions bio-informatiques en faveur d'un effet sur l'épissage (**Figure 20**), générant un taux de faux-positifs de 43%. Le variant *RAD51B* c.452+3A>G a déjà été décrit comme délétère dans une famille présentant des cas de cancers du sein et de l'ovaire (Golmard et al. 2013). Nos résultats basés sur les tests rapporteurs d'épissage ont montré que ce variant ne semble pas avoir d'effet sur l'épissage. Il conviendra d'obtenir un prélèvement sanguin afin d'étudier si cette absence d'effet est confirmée ou non. Par conséquent nous n'avons pu confirmer l'effet délétère de ce variant.

L'étude du profil d'épissage du gène *CHEK2* autour des exons 4 et 5 a montré l'existence d'un épissage alternatif correspondant au saut de l'exon 5 combiné à l'inclusion d'un exon cryptique de 359 nucléotides situé au début de l'intron 5 (r.574_721del; 721_722ins721+54_721+412). Parmi les 2 variants testés dans cette région, le c.449-5T>A est le plus fréquent de notre série (12 patients soit 0,38%) (**Tableau 2**) et a un effet partiel sur l'épissage (**Figures 20 et 21**), correspondant au saut combiné en phase des exons 4 et 5, exons codant un domaine fonctionnel protéique connu (FHA, *Forkhead-associated*). De manière intéressante, un patient porteur de ce variant (patient IV) est aussi porteur de la mutation délétère *BRCA1* c.68_69del provoquant l'introduction d'un codon STOP prématuré (p.Glu23Valfs*17). La contribution du variant d'épissage c.449-5T>A au risque global de cancer ne peut être écartée car ce variant est extrêmement rare dans la base de données ExAC

(8e-4). Son association avec une mutation délétère dans le gène *BRCA1* pourrait évoquer un modèle oligogénique.

Comme attendu, les tests fonctionnels basés sur l'utilisation de minigènes ont confirmé le défaut majeur d'épissage causé par le variant *RAD51C* c.1026+5_1026+7del, à savoir le saut total de l'exon 8 (r.966_1026del et p.Arg322Serfs*22) (**Figure 20**) (Golmard et al. 2013). De manière surprenante, cette mutation d'épissage a été détectée dans une famille ne présentant qu'un unique cas de cancer du sein (**Tableau 2**). Le variant *RAD51C* c.905-3C>G est une nouvelle mutation d'épissage, causant le renforcement total du saut hors phase de l'exon 7 (r.905_965del et p.Glu303Trpfs*41) (**Figures 20 et 21**). Cette mutation a été détectée dans une famille présentant des cas de cancers du sein et de l'ovaire (**Tableau 2**). Il n'y a pas encore eu d'études de co-ségrégation dans cette famille. Néanmoins, ces résultats suggèrent fortement que ce variant provoque la production d'une protéine non fonctionnelle, et sont en faveur d'une classification comme mutation pathogène. Les femmes porteuses de cette mutation auraient un risque plus élevé de développer un cancer de l'ovaire. Par conséquent cette mutation serait utilisable pour le conseil génétique.

Les mutations d'épissage introniques au niveau des sites consensus d'épissage des gènes liés au syndrome HBOC, autres que *BRCA1* et *BRCA2*, ne sont pas des événements rares et pourraient être utilisées pour le conseil génétique. Il n'a pas été observé de variation synonyme impactant potentiellement l'épissage. Il conviendrait d'étendre cette analyse à l'ensemble des types de variations exoniques dans ces régions, pouvant toucher des éléments régulateurs de type ESE ou ESS, mais également à l'ensemble des mutations pouvant créer des sites d'épissage, renforcer des sites cryptiques d'épissage ou encore abolir des sites de branchement. De plus, les recommandations, basées sur les prédictions bio-informatiques,

permettant de prioriser les études de l'impact de VSI sur l'épissage des gènes *BRCA1* et *BRCA2* ont été appliquées ici pour d'autres gènes. Nos résultats montrent que ces prédictions sont pertinentes pour d'autres gènes liés au syndrome HBOC mais qu'elles peuvent être améliorées. Ceci sera réalisé de manière prospective au laboratoire. Enfin, la technique de séquençage ciblé de l'ARN (RNA-Seq) permettrait, grâce à sa capacité de séquençage, d'analyser simultanément l'impact potentiel sur l'épissage de nombreux variants dans de nombreux gènes. Nous avons développé au laboratoire une telle approche et allons la tester de manière prospective. Par l'utilisation de cette technique, on pourrait envisager d'utiliser le test fonctionnel basé sur l'utilisation de minigènes comme test ciblé de confirmation d'un effet sur l'épissage, de même que la RT-PCR et le séquençage.

Ces travaux montrent que l'étude des VSI dans 9 autres gènes que *BRCA1* et *BRCA2* permet de détecter des mutations délétères d'épissage. Ces nouvelles mutations, pourront contribuer aux calculs de risques initiés dans l'étude Tumospec. Pour les variants d'interprétation délicate, de type effet partiel notamment, des études de co-ségrégation devront être initiées afin d'aider à leur classification.

Devant les nombreux VSI détectés dans le cadre du diagnostic moléculaire, la combinaison des approches *in silico* et *in vitro* apportent leur contribution à l'interprétation de l'impact de ces variants sur l'épissage. L'effort reste à faire pour la classification de ces variants selon leur pathogénicité, car les données fonctionnelles ne suffisent pas toujours. Il est indispensable de maitriser les concepts et les outils permettant d'interpréter correctement les VSI et de les classifier. Les bases de données sont un outil important pour le partage de l'interprétation de ces variants. Seulement celles-ci sont très nombreuses, redondantes voire contradictoires pour l'interprétation et la classification d'un même variant, ce qui peut être

dramatique pour les patients et leurs familles. Cela nécessite une expertise pluridisciplinaire, passant par des compétences en génétique moléculaire, bio-informatique, statistique, analyses fonctionnelles, mais aussi par la connaissance des données associées aux variants, comme les données de co-ségrégation, de co-occurrence, les données personnelles et familiales, les données tumorales et par la nécessité de collaborations et de partages tant au niveau régional que national et international, afin d'obtenir une puissance statistique suffisante. Un effort national, *via* le GGC, et international, *via* notamment les consortia Global Alliance ou ENIGMA, est réalisé, permettant d'augmenter régulièrement le nombre de mutations des gènes *BRCA1* et *BRCA2* utilisables pour le conseil génétique. Une réflexion est en cours pour initier le même effort pour les autres gènes impliqués dans la prédisposition aux cancers du sein et/ou de l'ovaire. En effet, aujourd'hui, l'interprétation et la classification correcte des variants est un enjeu crucial pour la prise en charge adaptée des patients et de leur famille, en particulier dans le cadre de la prédisposition héréditaire aux cancers du sein et de l'ovaire.

Références bibliographiques

Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broaddus R, Lu KH, Chen J, Tran TV, Williams D, et al. 2012. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer 107:1776–1782.

Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. 2015. Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. Hum Mutat 36:210–221.

Anczuków O, Buisson M, Léoné M, Coutanson C, Lasset C, Calender A, Sinilnikova OM, Mazoyer S. 2012. BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy. Clin Cancer Res Off J Am Assoc Cancer Res 18:4903–4909.

Anczuków O, Buisson M, Salles M-J, Triboulet S, Longy M, Lidereau R, Sinilnikova OM, Mazoyer S. 2008. Unclassified variants identified in BRCA1 exon 11: Consequences on splicing. Genes Chromosomes Cancer 47:418–426.

Anderson SL, Coli R, Daly IW, Kichula EA, Rork MJ, Volpi SA, Ekstein J, Rubin BY. 2001. Familial dysautonomia is caused by mutations of the IKAP gene. Am J Hum Genet 68:753–758.

Andreutti-Zaugg C, Scott RJ, Iggo R. 1997. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. Cancer Res 57:3288–3293.

Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, et al. 2014. Breast-cancer risk in families with mutations in PALB2. N Engl J Med 371:497–506.

Ast G. 2004. How did alternative splicing evolve? Nat Rev Genet 5:773–782.

Bai B, Hales CM, Chen P-C, Gozal Y, Dammer EB, Fritz JJ, Wang X, Xia Q, Duong DM, Street C, Cantero G, Cheng D, et al. 2013. U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. Proc Natl Acad Sci U S A 110:16562–16567.

Baralle D, Baralle M. 2005. Splicing in action: assessing disease causing sequence changes. J Med Genet 42:737–748.

Baralle D, Lucassen A, Buratti E. 2009. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep 10:810–816.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Çolak R, Kim T, Misquitta-Ali CM, et al. 2012. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science 338:1587–1593.

Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G. 2012. U1 snRNP determines mRNA length and regulates isoform expression. Cell 150:53–64.

Berget SM. 1995. Exon Recognition in Vertebrate Splicing. J Biol Chem 270:2411–2414.

Bernier FP, Caluseriu O, Ng S, Schwartzentruber J, Buckingham KJ, Innes AM, Jabs EW, Innis JW, Schuette JL, Gorski JL, Byers PH, Andelfinger G, et al. 2012. Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. Am J Hum Genet 90:925–933.

Béroud C, Letovsky SI, Braastad CD, Caputo SM, Beaudoux O, Bignon YJ, Bressac-De Paillerets B, Bronner M, Buell CM, Collod-Béroud G, Coulet F, Derive N, et al. 2016. BRCA Share: A Collection of Clinical BRCA Gene Variants. Hum Mutat 37:1318–1328.

Binder-Foucard F, Bossard N, Delafosse P, Belot A, Woronoff A-S, Remontet L, French network of cancer registries (Francim). 2014. Cancer incidence and mortality in France over the 1980-2012 period: solid tumors. Rev Dépidémiologie Santé Publique 62:95–108.

Birkbak NJ, Wang ZC, Kim J-Y, Eklund AC, Li Q, Tian R, Bowman-Colin C, Li Y, Greene-Colozzi A, Iglehart JD, Tung N, Ryan PD, et al. 2012. Telomeric Allelic Imbalance Indicates Defective DNA Repair and Sensitivity to DNA-Damaging Agents. ResearchGate 2:366–75.

Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res 21:5803–5816.

Blencowe BJ. 2006. Alternative Splicing: New Insights from Global Analyses. Cell 126:37–47.

Boise LH, González-García M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nuñez G, Thompson CB. 1993. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell 74:597–608.

Bonnet C, Krieger S, Vezain M, Rousselin A, Tournier I, Martins A, Berthet P, Chevrier A, Dugast C, Layet V, Rossi A, Lidereau R, et al. 2008. Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. J Med Genet 45:438–446.

Bork P, Blomberg N, Nilges M. 1996. Internal repeats in the BRCA2 protein sequence. Nat Genet 13:22–23.

Bouwman P, Gulden H van der, Heijden I van der, Drost R, Klijn CN, Prasetyanti P, Pieterse M, Wientjens E, Seibler J, Hogervorst FBL, Jonkers J. 2013. A High-Throughput Functional Complementation Assay for Classification of BRCA1 Missense Variants. Cancer Discov 3:1142–1155.

Broca P. 1866. Traité des tumeurs. P. Asselin. 624 p.

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. 2013. High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. J Med Genet 50:255–263.

Bunting SF, Callén E, Wong N, Chen H-T, Polato F, Gunn A, Bothmer A, Feldhahn N, Fernandez-Capetillo O, Cao L, Xu X, Deng C-X, et al. 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell 141:243–254.

Burge CB, Padgett RA, Sharp PA. 1998. Evolutionary Fates and Origins of U12-Type Introns. Mol Cell 2:773–785.

Byrnes GB, Southey MC, Hopper JL. 2008. Are the so-called low penetrance breast cancer genes, ATM, BRIP1, PALB2 and CHEK2, high risk for women with strong family histories? Breast Cancer Res BCR 10:208.

Camier S, Séraphin B. 2007. Détruisez ce message (ARN) après l'avoir lu ! médecine/sciences 23:850–856.

Caputo S, Benboudjema L, Sinilnikova O, Rouleau E, Béroud C, Lidereau R, French BRCA GGC Consortium. 2012. Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. Nucleic Acids Res 40:D992-1002.

Carmel I, Tal S, Vig I, Ast G. 2004. Comparative analysis detects dependencies among the 5' splice-site positions. RNA N Y N 10:828–840.

Carr K. 1993. Nobel goes to discoverers of "split genes." Nature 365:597.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298.

Cartegni L, Krainer AR. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 30:377–384.

Castéra L, Krieger S, Rousselin A, Legros A, Baumann J-J, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, Baert-Desurmont S, Tinat J, et al. 2014. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. Eur J Hum Genet EJHG 22:1305–1313.

Chang S, Biswas K, Martin BK, Stauffer S, Sharan SK. 2009. Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. J Clin Invest 119:3160–3171.

Chapman MS, Verma IM. 1996. Transcriptional activation by BRCA1. Nature 382:678–679.

Charlet-B N, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA. 2002. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. Mol Cell 10:45–53.

Chen PL, Chen CF, Chen Y, Xiao J, Sharp ZD, Lee WH. 1998. The BRC repeats in BRCA2 are critical for RAD51 binding and resistance to methyl methanesulfonate treatment. Proc Natl Acad Sci U S A 95:5287–5292.

Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, Deffenbaugh AM, Burbidge LA, Pruss D, Judkins T, Scholl T, Bekessy A, et al. 2006. Genetic and Histopathologic Evaluation of BRCA1 and BRCA2 DNA Sequence Variants of Unknown Clinical Significance. Cancer Res 66:2019–2027.

Chodosh LA. 1998. Expression of BRCA1 and BRCA2 in normal and neoplastic cells. J Mammary Gland Biol Neoplasia 3:389–402.

Claus EB, Risch N, Thompson WD. 1991. Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48:232–242.

Coene ED, Hollinshead MS, Waeytens AAT, Schelfhout VRJ, Eechaute WP, Shaw MK, Van Oostveldt PMV, Vaux DJ. 2005. Phosphorylated BRCA1 is predominantly located in the nucleus and mitochondria. Mol Biol Cell 16:997–1010.

Colombo M, Blok MJ, Whiley P, Santamariña M, Gutiérrez-Enríquez S, Romero A, Garre P, Becker A, Smith LD, De Vecchi G, Brandão RD, Tserpelis D, et al. 2014. Comprehensive annotation of splice

junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. Hum Mol Genet 23:3666–3680.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and Disease. Cell 136:777–793.

Cortez D, Wang Y, Qin J, Elledge SJ. 1999. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. Science 286:1162–1166.

Cuajungco MP, Leyne M, Mull J, Gill SP, Lu W, Zagzag D, Axelrod FB, Maayan C, Gusella JF, Slaugenhaupt SA. 2003. Tissue-specific reduction in splicing efficiency of IKBKAP due to the major mutation associated with familial dysautonomia. Am J Hum Genet 72:749–758.

Daguenet E, Dujardin G, Valcárcel J. 2015. The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO Rep 16:1640–1655.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA 4:49–60.

Deng CX, Brodie SG. 2000. Roles of BRCA1 and its interacting proteins. BioEssays News Rev Mol Cell Dev Biol 22:728–737.

Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67.

Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frébourg T, Tosi M, Martins A. 2013. Functional Analysis of a Large set of BRCA2 exon 7 Variants Highlights the Predictive Value of Hexamer Scores in Detecting Alterations of Exonic Splicing Regulatory Elements. Hum Mutat 34:1547–1557.

Dietrich RC, Incorvaia R, Padgett RA. 1997. Terminal Intron Dinucleotide Sequences Do Not Distinguish between U2- and U12-Dependent Introns. Mol Cell 1:151–160.

Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol 3:195–205.

Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, Tavtigian SV, Monteiro ANA, Iversen ES, Couch FJ, Goldgar DE. 2007. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am J Hum Genet 81:873–883.

Edery P, Marcaillou C, Sahbatou M, Labalme A, Chastang J, Touraine R, Tubacher E, Senni F, Bober MB, Nampoothiri S, Jouk P-S, Steichen E, et al. 2011. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. Science 332:240–243.

Eisinger F, Bressac B, Castaigne D, Cottu P-H, Lansac J, Lefranc J-P, Lesur A, Noguès C, Pierret J, Puy-Pernias S, Sobol H, Tardivon A, et al. 2006. [Identification and management of hereditary breast-ovarian cancers (2004 update)]. Pathol Biol (Paris) 54:230–250.

ElShamy WM, Livingston DM. 2004. Identification of BRCA1-IRIS, a BRCA1 locus product. Nat Cell Biol 6:954–967.

Esashi F, Christ N, Gannon J, Liu Y, Hunt T, Jasin M, West SC. 2005. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature 434:598–604.

Fackenthal JD, Yoshimatsu T, Zhang B, Garibay GR de, Colombo M, De Vecchi G, Ayoub SC, Lal K, Olopade OI, Vega A, Santamariña M, Blanco A, et al. 2016. Naturally occurring BRCA2 alternative mRNA splicing events in clinically relevant samples. J Med Genet.

Fairbrother WG, Yeh R-F, Sharp PA, Burge CB. 2002. Predictive Identification of Exonic Splicing Enhancers in Human Genes. Science 297:1007–1013.

Farrugia DJ, Agarwal MK, Pankratz VS, Deffenbaugh AM, Pruss D, Frye C, Wadum L, Johnson K, Mentlick J, Tavtigian SV, Goldgar DE, Couch FJ. 2008. Functional Assays for Classification of BRCA2 Variants of Uncertain Significance. Cancer Res 68:3523–3531.

Fernández-Nogales M, Cabrera JR, Santos-Galindo M, Hoozemans JJM, Ferrer I, Rozemuller AJM, Hernández F, Avila J, Lucas JJ. 2014. Huntington's disease is a four-repeat tauopathy with tau nuclear rods. Nat Med 20:881–885.

Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, Dunnen JT den. 2011. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat 32:557–563.

Foray N, Marot D, Gabriel A, Randrianarison V, Carr AM, Perricaudet M, Ashworth A, Jeggo P. 2003. A subset of ATM- and ATR-dependent phosphorylation events requires the BRCA1 protein. EMBO J 22:2860–2871.

Fortin J, Moisan A-M, Dumont M, Leblanc G, Labrie Y, Durocher F, Bessette P, Bridge P, Chiquette J, Laframboise R, Lépine J, Lespérance B, et al. 2005. A new alternative splice variant of BRCA1 containing an additional in-frame exon. Biochim Biophys Acta 1731:57–65.

Fox-Walsh KL, Dou Y, Lam BJ, Hung S, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 102:16176–16181.

Frankish A, Mudge JM, Thomas M, Harrow J. 2012. The importance of identifying alternative splicing in vertebrate genome annotation. Database 2012:bas014.

Frebourg T. 2014. The challenge for the next generation of medical geneticists. Hum Mutat 35:909–911.

Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG. 2015. RNA-Binding Proteins: Splicing Factors and Disease. Biomolecules 5:893–909.

Frilander MJ, Steitz JA. 1999. Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. Genes Dev 13:851–863.

Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. RNA 1:663–680.

Fu X-D, Ares M. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 15:689–701.

Fuks F, Milner J, Kouzarides T. 1998. BRCA2 associates with acetyltransferase activity when bound to P/CAF. Oncogene 17:2531–2534.

Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. 2010a. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol Clifton NJ 653:249–257.

Gaildrat P, Krieger S, Di Giacomo D, Abdat J, Révillion F, Caputo S, Vaur D, Jamard E, Bohers E, Ledemeney D, Peyrat J-P, Houdayer C, et al. 2012. Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. J Med Genet 49:609–617.

Gaildrat P, Krieger S, Théry J-C, Killian A, Rousselin A, Berthet P, Frébourg T, Hardouin A, Martins A, Tosi M. 2010b. The BRCA1 c.5434C->G (p.Pro1812Ala) variant induces a deleterious exon 23 skipping by affecting exonic splicing regulatory elements. J Med Genet 47:398–403.

Gerbino V, Carrì MT, Cozzolino M, Achsel T. 2013. Mislocalised FUS mutants stall spliceosomal snRNPs in the cytoplasm. Neurobiol Dis 55:120–128.

Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, Cruz-Correa M, Offerhaus JA. 2000. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 119:1447–1453.

Gilbert W. 1978. Why genes in pieces? Nature 271:501.

Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K, Lührmann R. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat Commun 3:994.

Global Alliance for Genomics and Health. 2016. GENOMICS. A federated ecosystem for sharing genomic, clinical data. Science 352:1278–1280.

Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ, Breast Cancer Information Core (BIC) Steering Committee. 2004. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am J Hum Genet 75:535–544.

Golmard L, Caux-Moncoutier V, Davy G, Al Ageeli E, Poirot B, Tirapo C, Michaux D, Barbaroux C, Enghien CD d', Nicolas A, Castéra L, Sastre-Garau X, et al. 2013. Germline mutation in the RAD51B gene confers predisposition to breast cancer. BMC Cancer 13:484.

Gooding C, Clark F, Wollerton MC, Grellscheid S-N, Groom H, Smith CWJ. 2006. A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. Genome Biol 7:R1.

Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, Pupko T, Ast G. 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. Mol Cell 22:769–781.

Gowen LC, Johnson BL, Latour AM, Sulik KK, Koller BH. 1996. Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. Nat Genet 12:191–194.

Gozani O, Feld R, Reed R. 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev 10:233–243.

Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol 18:4752–4760.

Graveley BR. 2001. Alternative splicing: increasing diversity in the proteomic world. Trends Genet TIG 17:100–107.

Greenberg RA. 2008. Recognition of DNA double strand breaks by the BRCA1 tumor suppressor network. Chromosoma 117:305–317.

Groen EJN, Fumoto K, Blokhuis AM, Engelen-Lee J, Zhou Y, Heuvel DMA van den, Koppers M, Diggelen F van, Heest J van, Demmers JAA, Kirby J, Shaw PJ, et al. 2013. ALS-associated mutations in FUS disrupt the axonal distribution and function of SMN. Hum Mol Genet 22:3690–3704.

Grover A, Houlden H, Baker M, Adamson J, Lewis J, Prihar G, Pickering-Brown S, Duff K, Hutton M. 1999. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. J Biol Chem 274:15134–15143.

Gui JF, Lane WS, Fu XD. 1994. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature 369:678–682.

Guidugli L, Carreira A, Caputo SM, Ehlen A, Galli A, Monteiro ANA, Neuhausen SL, Hansen TVO, Couch FJ, Vreeswijk MPG, on behalf of the ENIGMA consortium. 2014. Functional Assays for Analysis of Variants of Uncertain Significance in BRCA2. Hum Mutat 35:151–164.

Guidugli L, Pankratz VS, Singh N, Thompson J, Erding CA, Engel C, Schmutzler R, Domchek S, Nathanson K, Radice P, Singer C, Tonin PN, et al. 2013. A Classification Model for BRCA2 DNA Binding Domain Missense Variants Based on Homology-Directed Repair Activity. Cancer Res 73:265–275.

Hakem R, Pompa JL de la, Sirard C, Mo R, Woo M, Hakem A, Wakeham A, Potter J, Reitmair A, Billia F, Firpo E, Hui CC, et al. 1996. The tumor suppressor gene Brca1 is required for embryonic cellular proliferation in the mouse. Cell 85:1009–1023.

He H, Liyanarachchi S, Akagi K, Nagy R, Li J, Dietrich RC, Li W, Sebastian N, Wen B, Xin B, Singh J, Yan P, et al. 2011. Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science 332:238–240.

Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB. 2008. Tissue-Specific Genetic Control of Splicing: Implications for the Study of Complex Traits. PLOS Biol 6:e1000001.

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. 2012. Rare variants in XRCC2 as breast cancer susceptibility alleles. J Med Genet 49:618–620.

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59.

Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Léone M, et al. 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. Hum Mutat 33:1228–1238.

Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pagès-Berhouet S, Enghien CD d', Laugé A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D. 2008. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat 29:975–982.

Howard JM, Sanford JR. 2015. THE RNAissance Family: SR proteins as multifaceted regulators of gene expression. Wiley Interdiscip Rev RNA 6:93–110.

Hsu LC, White RL. 1998. BRCA1 is associated with the centrosome during mitosis. Proc Natl Acad Sci U S A 95:12983–12988.

Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin S-F, Milner J, Brown LA, Hsu F, Gilks B, Nielsen T, Schulzer M, et al. 2003. EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell 115:523–535.

Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, et al. 1998. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393:702–705.

Ibrahim EC, Hims MM, Shomron N, Burge CB, Slaugenhaupt SA, Reed R. 2007. Weak definition of IKBKAP exon 20 leads to aberrant splicing in familial dysautonomia. Hum Mutat 28:41–53.

Institut National du Cancer. 2016.

Ishihara T, Ariizumi Y, Shiga A, Kato T, Tan C-F, Sato T, Miki Y, Yokoo M, Fujino T, Koyama A, Yokoseki A, Nishizawa M, et al. 2013. Decreased number of Gemini of coiled bodies and U12 snRNA level in amyotrophic lateral sclerosis. Hum Mol Genet 22:4136–4147.

Izquierdo JM, Majós N, Bonnal S, Martínez C, Castelo R, Guigó R, Bilbao D, Valcárcel J. 2005. Regulation of Fas Alternative Splicing by Antagonistic Effects of TIA-1 and PTB on Exon Definition. Mol Cell 19:475–484.

Jensen CJ, Oldfield BJ, Rubio JP. 2009. Splicing, cis genetic variation and disease. Biochem Soc Trans 37:1311–1315.

Kais Z, Chiba N, Ishioka C, Parvin JD. 2012. Functional differences among BRCA1 missense mutations in the control of centrosome duplication. Oncogene 31:799–804.

Kalsotra A, Cooper TA. 2011. Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet 12:715–729.

Kashima T, Manley JL. 2003. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nat Genet 34:460–463.

Kast K, Rhiem K, Wappenschmidt B, Hahnen E, Hauke J, Bluemcke B, Zarghooni V, Herold N, Ditsch N, Kiechle M, Braun M, Fischer C, et al. 2016. Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. J Med Genet 53:465–471.

Kfir N, Lev-Maor G, Glaich O, Alajem A, Datta A, Sze SK, Meshorer E, Ast G. 2015. SF3B1 association with chromatin determines splicing outcomes. Cell Rep 11:618–629.

Kim E, Magen A, Ast G. 2007. Different levels of alternative splicing among eukaryotes. Nucleic Acids Res 35:125–131.

Kobayashi H, Ohno S, Sasaki Y, Matsuura M. 2013. Hereditary breast and ovarian cancer susceptibility genes (review). Oncol Rep 30:1019–1029.

Konarska MM, Vilardell J, Query CC. 2006. Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. Mol Cell 21:543–553.

Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN. 2007. Single base-pair substitutions in exon–intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat 28:150–158.

Krecic AM, Swanson MS. 1999. hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol 11:363–371.

Kugler W, Enssle J, Hentze MW, Kulozik AE. 1995. Nuclear degradation of nonsense mutated β-globin mRNA: a post-transcriptional mechanism to protect heterozygotes from severe clinical manifestations of β-thalassemia? Nucleic Acids Res 23:413–418.

Kurosaki T, Maquat LE. 2016. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci 129:461–467.

Kuyumcu-Martinez NM, Wang G-S, Cooper TA. 2007. Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Mol Cell 28:68–78.

Kuznetsov SG, Liu P, Sharan SK. 2008. Mouse embryonic stem cell–based functional assay to evaluate mutations in BRCA2. Nat Med 14:875–881.

Kwiatkowski TJ, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T, Valdmanis P, Rouleau GA, et al. 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323:1205–1208.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980-985.

Lee J-H, Paull TT. 2005. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science 308:551–554.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem 84:291–323.

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M. 1995. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 80:155–165.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Lewandowska MA. 2013. The missing puzzle piece: splicing mutations. Int J Clin Exp Pathol 6:2675–2682.

Lin X, Miller JW, Mankodi A, Kanadia RN, Yuan Y, Moxley RT, Swanson MS, Thornton CA. 2006. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. Hum Mol Genet 15:2087–2097.

Lindor NM, Guidugli L, Wang X, Vallée MP, Monteiro ANA, Tavtigian S, Goldgar DE, Couch FJ. 2012. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). Hum Mutat 33:8–21.

Lines MA, Huang L, Schwartzentruber J, Douglas SL, Lynch DC, Beaulieu C, Guion-Almeida ML, Zechi-Ceide RM, Gener B, Gillessen-Kaesbach G, Nava C, Baujat G, et al. 2012. Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am J Hum Genet 90:369–377.

Lingle WL, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL. 1998. Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. Proc Natl Acad Sci U S A 95:2950–2955.

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 293:864–867.

Lorson CL, Hahnen E, Androphy EJ, Wirth B. 1999. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci U S A 96:6307–6311.

Loveday C, Turnbull C, Ruark E, Xicola RMM, Ramsay E, Hughes D, Warren-Perry M, Snape K, Breast Cancer Susceptibility Collaboration (BCSC) (uk), Eccles D, Evans DG, Gore M, et al. 2012. Germline RAD51C mutations confer susceptibility to ovarian cancer. Nat Genet 44:475–476.

Lynch HT, Guirgis HA, Albert S, Brennan M, Lynch J, Kraft C, Pocekay D, Vaughns C, Kaplan A. 1974. Familial association of carcinoma of the breast and ovary. Surg Gynecol Obstet 138:717–724.

Lynch HT, Harris RE, Guirgis HA, Lynch PM, Maloney K, Rankin L, Lynch J. 1976. Early age of onset and familial breast cancer. Lancet Lond Engl 2:626–627.

Lynch HT, Krush AJ. 1971. Carcinoma of the breast and ovary in three families. Surg Gynecol Obstet 133:644–648.

MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, et al. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983.

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barceló J, O'Hoy K. 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255:1253–1255.

Manke IA, Lowery DM, Nguyen A, Yaffe MB. 2003. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science 302:636–639.

Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, Cannon SC, Thornton CA. 2002. Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell 10:35–44.

Manley JL, Krainer AR. 2010. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes Dev 24:1073–1074.

Martinez JS, Baldeyron C, Carreira A. 2015. Molding BRCA2 function through its interacting partners. Cell Cycle Georget Tex 14:3389–3395.

Matera AG, Wang Z. 2014. A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15:108–121.

Mavaddat N, Peock S, Frost D, Ellis S, Platte R, Fineberg E, Evans DG, Izatt L, Eeles RA, Adlard J, Davidson R, Eccles D, et al. 2013. Cancer Risks for BRCA1 and BRCA2 Mutation Carriers: Results From Prospective Analysis of EMBRACE. J Natl Cancer Inst 105:812–822.

McCullough AJ, Berget SM. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17:4562–4571.

McCullough AJ, Berget SM. 2000. An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites. Mol Cell Biol 20:9225–9235.

McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, Understanding Society Scientific Group, Crawley C, Craig J, Scott MA, Hodkinson C, Baxter J, et al. 2015. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep 10:1239–1245.

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. 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet 42:410–414.

Michael D, Manyuan L. 1999. Intron—exon structures of eukaryotic model organisms. Nucleic Acids Res 27:3219–3228.

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

Millot GA, Carvalho MA, Caputo SM, Vreeswijk MPG, Brown MA, Webb M, Rouleau E, Neuhausen SL, Hansen T v. O, Galli A, Brandão RD, Blok MJ, et al. 2012. A guide for functional analysis of BRCA1 variants of uncertain significance. Hum Mutat 33:1526–1537.

Milner J, Ponder B, Hughes-Davies L, Seltmann M, Kouzarides T. 1997. Transcriptional activation functions in BRCA2. Nature 386:772–773.

Minovitsky S, Gee SL, Schokrpur S, Dubchak I, Conboy JG. 2005. The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res 33:714–724.

Monteiro AN, August A, Hanafusa H. 1996. Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc Natl Acad Sci U S A 93:13595–13599.

Mordes D, Luo X, Kar A, Kuo D, Xu L, Fushimi K, Yu G, Sternberg P, Wu JY. 2006. Pre-mRNA splicing and retinitis pigmentosa. Mol Vis 12:1259–1271.

Mort M, Sterne-Weiler T, Li B, Ball EV, Cooper DN, Radivojac P, Sanford JR, Mooney SD. 2014. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol 15:R19.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628.

Mudge JM, Frankish A, Fernandez-Banet J, Alioto T, Derrien T, Howald C, Reymond A, Guigó R, Hubbard T, Harrow J. 2011. The Origins, Evolution, and Functional Potential of Alternative Splicing in Vertebrates. Mol Biol Evol 28:2949–2959.

Muller D, Rouleau E, Schultz I, Caputo S, Lefol C, Bièche I, Caron O, Noguès C, Limacher JM, Demange L, Lidereau R, Fricker JP, et al. 2011. An entire exon 3 germ-line rearrangement in the BRCA2 gene: pathogenic relevance of exon 3 deletion in breast cancer predisposition. BMC Med Genet 12:121.

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci 23:198–199.

Neidhardt G, Becker A, Hauke J, Horvath J, Bogdanova Markov N, Heilmann-Heimbach S, Hellebrand H, Thiele H, Altmuller J, Nurnberg P, Meindl A, Rhiem K, et al. 2016. The RAD51C exonic splice-site mutations c.404G>C and c.404G>T are associated with familial breast and ovarian cancer. J Cancer Prev.

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, et al. 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314:130–133.

Orban TI, Olah E. 2003. Emerging roles of BRCA1 alternative splicing. Mol Pathol 56:191–197.

Osorio A, Endt D, Fernández F, Eirich K, Hoya M de la, Schmutzler R, Caldés T, Meindl A, Schindler D, Benitez J. 2012. Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. Hum Mol Genet 21:2889–2898.

Ouchi T. 2006. BRCA1 phosphorylation: biological consequences. Cancer Biol Ther 5:470–475.

Padgett RA. 2012. New connections between splicing and human disease. Trends Genet 28:147–154.

Pagani F, Baralle FE. 2004. Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389–396.

Pagani F, Raponi M, Baralle FE. 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A 102:6368–6372.

Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, Pellagatti A, Wainscoat JS, Hellstrom-Lindberg E, Gambacorti-Passerini C, Godfrey AL, Rapado I, et al. 2011. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 365:1384–1395.

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. 2012. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet 90:734–739.

Patel AA, Steitz JA. 2003. Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol 4:960–970.

Pharoah PDP, Guilford P, Caldas C. 2001. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 121:1348–1353.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian SV, IARC Unclassified Genetic Variants Working Group. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282–1291.

Popova T, Manié 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. 2012. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with BRCA1/2 Inactivation. Cancer Res 72:5454–5462.

Pottier C, Ravenscroft TA, Sanchez-Contreras M, Rademakers R. 2016. Genetics of FTLD: overview and what else we can expect from genetic studies. J Neurochem 138:32–53.

Pozzoli U, Sironi M. 2005. Silencers regulate both constitutive and alternative splicing events in mammals. Cell Mol Life Sci CMLS 62:1579–1604.

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

Puig O, Gottschalk A, Fabrizio P, Séraphin B. 1999. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. Genes Dev 13:569–580.

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, et al. 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44:47–52.

Ransburgh DJR, Chiba N, Ishioka C, Toland AE, Parvin JD. 2010. Identification of Breast Tumor Mutations in BRCA1 That Abolish Its Function in Homologous DNA Recombination. Cancer Res 70:988–995.

Raponi M, Douglas AGL, Tammaro C, Wilson DI, Baralle D. 2012. Evolutionary constraint helps unmask a splicing regulatory region in BRCA1 exon 11. PloS One 7:e37255.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-Scale Proteomic Analysis of the Human Spliceosome. Genome Res 12:1231–1245.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol J Comput Mol Cell Biol 4:311–323.

Renton AE, Chiò A, Traynor BJ. 2014. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 17:17–23.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE, Molecular Subcommittee of the ACMG Laboratory Quality Assurance Committee. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med Off J Am Coll Med Genet 10:294–300.

Rodríguez JA, Henderson BR. 2000. Identification of a functional nuclear export sequence in BRCA1. J Biol Chem 275:38589–38596.

Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, Guarini A, Del Giudice I, et al. 2011. Mutations of the SF3B1 splicing factor in chronic

lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood 118:6904–6908.

Roy R, Chun J, Powell SN. 2012. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer 12:68–78.

Sakabe NJ, Souza SJ de. 2007. Sequence features responsible for intron retention in human. BMC Genomics 8:59.

Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RLM, Roos RAC, Howland D, Detloff PJ, Housman DE, et al. 2013. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A 110:2366–2370.

Savage KI, Harkin DP. 2015. BRCA1, a “complex” protein involved in the maintenance of genomic stability. FEBS J 282:630–646.

Savkur RS, Philips AV, Cooper TA. 2001. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. Nat Genet 29:40–47.

Schneider C, Will CL, Makarova OV, Makarov EM, Lührmann R. 2002. Human U4/U6.U5 and U4atac/U6atac.U5 Tri-snRNPs Exhibit Similar Protein Compositions. Mol Cell Biol 22:3219–3229.

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. 2013. RAD51C deletion screening identifies a recurrent gross deletion in breast cancer and ovarian cancer families. Breast Cancer Res 15:R120.

Serra E, Ars E, Ravella A, Sánchez A, Puig S, Rosenbaum T, Estivill X, Lázaro C. 2001. Somatic NF1 mutational spectrum in benign neurofibromas: mRNA splice defects are common among point mutations. Hum Genet 108:416–429.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15:7155–7174.

Sharan SK, Morimatsu M, Albrecht U, Lim DS, Regel E, Dinh C, Sands A, Eichele G, Hasty P, Bradley A. 1997. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. Nature 386:804–810.

Shchepachev V, Wischnewski H, Missiaglia E, Soneson C, Azzalin CM. 2012. Mpn1, mutated in poikiloderma with neutropenia protein 1, is a conserved 3’-to-5’ RNA exonuclease processing U6 small nuclear RNA. Cell Rep 2:855–865.

Shen D, Vadgama JV. 1999. BRCA1 and BRCA2 gene mutation analysis: visit to the Breast Cancer Information Core (BIC). Oncol Res 11:63–69.

Shen H, Green MR. 2004. A Pathway of Sequential Arginine-Serine-Rich Domain-Splicing Signal Interactions during Mammalian Spliceosome Assembly. Mol Cell 16:363–373.

Shen H, Green MR. 2007. RS domain–splicing signal interactions in splicing of U12-type and U2-type introns. Nat Struct Mol Biol 14:597–603.

Shen H, Kan JLC, Green MR. 2004. Arginine-Serine-Rich Domains Bound at Splicing Enhancers Contact the Branchpoint to Promote Prespliceosome Assembly. Mol Cell 13:367–376.

Shimizu Y, Luk H, Horio D, Miron P, Griswold M, Iglehart D, Hernandez B, Killeen J, ElShamy WM. 2012. BRCA1-IRIS Overexpression Promotes Formation of Aggressive Breast Cancers. PLOS ONE 7:e34102.

Shin S, Verma IM. 2003. BRCA2 cooperates with histone acetyltransferases in androgen receptor-mediated transcription. Proc Natl Acad Sci U S A 100:7201–7206.

Singh RK, Cooper TA. 2012. Pre-mRNA splicing in disease and therapeutics. Trends Mol Med 18:472–482.

Smith EC. 2012. An overview of hereditary breast and ovarian cancer syndrome. J Midwifery Womens Health 57:577–584.

Sopik V, Akbari M r., Narod S a. 2015. Genetic testing for RAD51C mutations: in the clinic and community. Clin Genet 88:303–312.

Sorek R, Ast G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13:1631–1637.

Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. 2016. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLoS Genet 12:e1005756.

Spurdle AB, Couch FJ, Hogervorst FBL, Radice P, Sinilnikova OM, IARC Unclassified Genetic Variants Working Group. 2008. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 29:1304–1313.

Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, Radice P, Stoppa-Lyonnet D, Tavtigian S, Wappenschmidt B, Couch FJ, Goldgar DE, et al. 2012. ENIGMA—Evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat 33:2–7.

Staley JP, Guthrie C. 1998. Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things. Cell 92:315–326.

Steffensen AY, Dandanell M, Jønson L, Ejlertsen B, Gerdes A-M, Nielsen FC, Hansen T vO. 2014. Functional characterization of BRCA1 gene variants by mini-gene splicing assay. Eur J Hum Genet EJHG 22:1362–1368.

Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. 2011. Loss of exon identity is a common mechanism of human inherited disease. Genome Res 21:1563–1571.

Sterne-Weiler T, Sanford JR. 2014. Exon identity crisis: disease-causing mutations that disrupt the splicing code. Genome Biol 15:201.

Sugnet CW, Kent WJ, Ares M, Haussler D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. Pac Symp Biocomput Pac Symp Biocomput 66–77.

Sun S, Ling S-C, Qiu J, Albuquerque CP, Zhou Y, Tokunaga S, Li H, Qiu H, Bui A, Yeo GW, Huang EJ, Eggan K, et al. 2015. ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. Nat Commun 6:6171.

Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. Cell 156:1324–1335.

Sy SMH, Huen MSY, Chen J. 2009. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci 106:7155–7160.

Tai YC, Domchek S, Parmigiani G, Chen S. 2007. Breast Cancer Risk Among Male BRCA1 and BRCA2 Mutation Carriers. J Natl Cancer Inst 99:1811–1814.

Tammaro C, Raponi M, Wilson DI, Baralle D. 2012. BRCA1 exon 11 alternative splicing, multiple functions and the association with cancer. Biochem Soc Trans 40:768–772.

Tarn W-Y, Steitz JA. 1996. Highly Diverged U4 and U6 Small Nuclear RNAs Required for Splicing Rare AT-AC Introns. Science 273:1824–1832.

Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. 2008. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat 29:1342–1354.

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, Thorlacius S, Offit K, Stoppa-Lyonnet D, Belanger C, Bell R, et al. 1996. The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. Nat Genet 12:333–337.

Telli ML, Timms KM, Reid JE, Hennessy B, Mills GB, Jensen KC, Szallasi Z, Barry WT, Winer EP, Tung N, Isakoff SJ, Ryan PD, et al. 2016. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple Negative Breast Cancer. Clin Cancer Res clincanres.2477.2015.

Thakur S, Zhang HB, Peng Y, Le H, Carroll B, Ward T, Yao J, Farid LM, Couch FJ, Wilson RB, Weber BL. 1997. Localization of BRCA1 and a splice variant identifies the nuclear localization signal. Mol Cell Biol 17:444–452.

Théry JC, Krieger S, Gaildrat P, Révillion F, Buisine M-P, Killian A, Duponchel C, Rousselin A, Vaur D, Peyrat J-P, Berthet P, Frébourg T, et al. 2011. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. Eur J Hum Genet EJHG 19:1052–1058.

Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I, Capellá G, Dunnen JT den, Sart D du, Fabre A, et al. 2014. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 46:107–115.

Thorslund T, West SC. 2007. BRCA2: a universal recombinase regulator. Oncogene 26:7720–7730.

Turunen JJ, Niemelä EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA 4:61–76.

Vail PJ, Morris B, Kan A van, Burdett BC, Moyes K, Theisen A, Kerr ID, Wenstrup RJ, Eggington JM. 2015. Comparison of locus-specific databases for BRCA1 and BRCA2 variants reveals disparity in variant classification within and among databases. J Community Genet 6:351–359.

Valcárcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. Science 273:1706–1709.

Vallée MP, Di Sera TL, Nix DA, Paquette AM, Parsons MT, Bell R, Hoffman A, Hogervorst FBL, Goldgar DE, Spurdle AB, Tavtigian SV. 2016. Adding In Silico Assessment of Potential Splice Aberration to the Integrated Evaluation of BRCA Gene Unclassified Variants. Hum Mutat 37:627–639.

Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, et al. 2009. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science 323:1208–1211.

Venables JP. 2004. Aberrant and alternative splicing in cancer. Cancer Res 64:7647–7654.

Venkitaraman AR. 2009. Linking the cellular functions of BRCA genes to cancer pathogenesis and treatment. Annu Rev Pathol 4:461–487.

Volpi L, Roversi G, Colombo EA, Leijsten N, Concolino D, Calabria A, Mencarelli MA, Fimiani M, Macciardi F, Pfundt R, Schoenmakers EFPM, Larizza L. 2010. Targeted next-generation sequencing appoints c16orf57 as clericuzio-type poikiloderma with neutropenia gene. Am J Hum Genet 86:72–76.

Wahl MC, Will CL, Lührmann R. 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136:701–718.

Walker LC, Whiley PJ, Houdayer C, Hansen TVO, Vega A, Santamarina M, Blanco A, Fachal L, Southey MC, Lafferty A, Colombo M, De Vecchi G, et al. 2013. Evaluation of a 5-tier scheme proposed for classification of sequence variants using bioinformatic and splicing assay data: inter-reviewer variability and promotion of minimum reporting guidelines. Hum Mutat 34:1424–1431.

Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, Roach KC, Mandell J, Lee MK, Ciernikova S, Foretova L, Soucek P, et al. 2006. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA 295:1379–1388.

Wang C, Chua K, Seghezzi W, Lees E, Gozani O, Reed R. 1998. Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev 12:1409–1414.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 8:749–761.

Wang J, Smith PJ, Krainer AR, Zhang MQ. 2005. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. Nucleic Acids Res 33:5053–5062.

Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA N Y N 14:802–813.

Wang Z, Xiao X, Van Nostrand E, Burge CB. 2006. General and Specific Functions of Exonic Splicing Silencers in Splicing Control. Mol Cell 23:61–70.

Ward AJ, Cooper TA. 2010. The pathobiology of splicing. J Pathol 220:152–163.

Wassarman KM, Steitz JA. 1992. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. Mol Cell Biol 12:1276–1285.

West SC. 2003. Molecular views of recombination proteins and their control. Nat Rev Mol Cell Biol 4:435–445.

Whiley PJ, Hoya M de la, Thomassen M, Becker A, Brandão R, Pedersen IS, Montagna M, Menéndez M, Quiles F, Gutiérrez-Enríquez S, Leeneer KD, Tenés A, et al. 2014. Comparison of mRNA Splicing Assay Protocols across Multiple Laboratories: Recommendations for Best Practice in Standardized Clinical Testing. Clin Chem 60:341–352.

Will CL, Schneider C, Hossbach M, Urlaub H, Rauhut R, Elbashir S, Tuschl T, Lührmann R. 2004. The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA 10:929–941.

Wong AK, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. 1997. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. J Biol Chem 272:31941–31944.

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

Wu J, Li Y, Jiang R. 2014. Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies. PLoS Genet 10:e1004237.

Wu W, Koike A, Takeshita T, Ohta T. 2008. The ubiquitin E3 ligase activity of BRCA1 and its biological functions. Cell Div 3:1.

Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ, Livingston DM. 2006. Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. Mol Cell 22:719–729.

Xia Y, Pao GM, Chen H-W, Verma IM, Hunter T. 2003. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. J Biol Chem 278:5255–5263.

Yamazaki T, Chen S, Yu Y, Yan B, Haertlein TC, Carrasco MA, Tapia JC, Zhai B, Das R, Lalancette-Hebert M, Sharma A, Chandran S, et al. 2012. FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. Cell Rep 2:799–806.

Yano K, Morotomi K, Saito H, Kato M, Matsuo F, Miki Y. 2000. Nuclear localization signals of the BRCA2 protein. Biochem Biophys Res Commun 270:171–175.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol J Comput Mol Cell Biol 11:377–394.

Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478:64–69.

Yu X, Chini CCS, He M, Mer G, Chen J. 2003. The BRCT domain is a phospho-protein binding domain. Science 302:639–642.

Yu X, Wu LC, Bowcock AM, Aronheim A, Baer R. 1998. The C-terminal (BRCT) Domains of BRCA1 Interact in Vivo with CtIP, a Protein Implicated in the CtBP Pathway of Transcriptional Repression. J Biol Chem 273:25388–25392.

Yu Y, Chi B, Xia W, Gangopadhyay J, Yamazaki T, Winkelbauer-Hurt ME, Yin S, Eliasse Y, Adams E, Shaw CE, Reed R. 2015. U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. Nucleic Acids Res 43:3208–3218.

Yuan SS, Lee SY, Chen G, Song M, Tomlinson GE, Lee EY. 1999. BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo. Cancer Res 59:3547–3551.

Zhang D, Rosbash M. 1999. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. Genes Dev 13:581–592.

Zhang F, Fan Q, Ren K, Andreassen PR. 2009. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol Cancer Res MCR 7:1110–1118.

Zhang J, Powell SN. 2005. The role of the BRCA1 tumor suppressor in DNA double-strand break repair. Mol Cancer Res MCR 3:531–539.

Zhang MQ. 1998. Statistical Features of Human Exons and Their Flanking Regions. Hum Mol Genet 7:919–932.

Zhou Z, Fu X-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma 122:191–207.

Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. Nature 419:182–185.

Annexes

Annexe I : Liste des amorces utilisées pour les tests fonctionnels basés sur l'utilisation de minigènes.

Gène	**Amorce sens**	**Amorce antisens**	**Taille de l'amplicon**
BRIP1 Exon 17	GCTTGAGATCACACAGCTGA	GCATCTTTGTGTGCTATTCATTG	600 pb
CHEK2 Exons 4-5	TCTTGGTAGAGATGGGTGCC	CAATGCAACTAGGACGGCAA	1 183 pb
RAD51B Exon 5	GCCAATTCAGATGCAGAGGG	CATGAGCCCAGATCGTGC	509 pb
RAD51B Exon 10	AAGGTGACCACTTTGAAAGGA	AATTCTTGCCACAACTGCCG	445 pb
RAD51C Exon 7	ACGTTAGGCTTGATGATATGGA	ACTCATTTCAGGTGCATGCG	582 pb
RAD51C Exon 8	ACGGGTAATTTGAAGGGTGT	AAGTGGGCTATATTTGGTGCT	612 pb

Annexe II : Liste des amorces utilisées pour la réaction de RT-PCR et la réaction de séquence lors des tests fonctionnels basés sur l'utilisation de minigènes.

Nom de l'amorce	Séquence
pCAS-Seq-F	GGGGTCAATAGCAGTGAGAG
pCAS-KO1-F	TGACGTCGCCGCCCATCAC
pCAS-2-R	ATTGGTTGTTGAGTTGGTTGTC

Annexe III : Liste des amorces utilisées pour la réaction de RT-PCR et la réaction de séquence.

Gène	**Amorce sens**	**Amorce antisens**	**Taille de l'amplicon**
CHEK2 Exons 3-6	GGGATAAGAGGTGTGAGGCA	TGCCTTAGGATAAACTGACTGAT	372 pb
RAD51C Exons 6-8	AACCAATCAGATGACAACAAAGA	CTTCTGGCTGGGTGACTTGT	153 pb

VU, LE PRESIDENT DU JURY

CAEN, LE

VU, LE DIRECTEUR DE L'UFR

CAEN, LE

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses et mémoires. Ces opinions doivent être considérées comme propres à leurs auteurs.