



HAL
open science

Caractérisation fonctionnelle de variations splicéogéniques à l'origine d'anomalies d'épissage en phase dans des gènes de prédisposition aux cancers : implications en génétique médicale

Laetitia Meulemans

► **To cite this version:**

Laetitia Meulemans. Caractérisation fonctionnelle de variations splicéogéniques à l'origine d'anomalies d'épissage en phase dans des gènes de prédisposition aux cancers : implications en génétique médicale. Génétique humaine. Normandie Université, 2021. Français. NNT : 2021NORMR008 . tel-03215013

HAL Id: tel-03215013

<https://theses.hal.science/tel-03215013>

Submitted on 3 May 2021

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

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



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

**Spécialité Aspects moléculaires et cellulaires de la biologie
- Génétique du cancer -**

Préparée au sein de l'Université de Rouen Normandie

Caractérisation fonctionnelle de variations splicéogéniques à l'origine d'anomalies d'épissage en phase dans des gènes de prédisposition aux cancers : implications en génétique médicale

**Présentée et soutenue par
Laëtitia MEULEMANS**

**Thèse soutenue publiquement le 31 mars 2021
devant le jury composé de**

Mme Béatrice PARFAIT	MCU-PH, AP-HP Hôpital Cochin	Rapporteur
M. Claude FEREC	PU-PH, CHRU de Brest Hôpital Morvan	Rapporteur
Mme Florence COULET	MCU-PH, AP-HP Pitié Salpêtrière	Examineur
M. David TULASNE	DR, Institut Pasteur de Lille	Examineur
M. Vincent RICHARD	PU-PH, Université de Rouen Normandie	Examineur
M. Claude HOUDAYER	PU-PH, Université de Rouen Normandie	Examineur
Mme Alexandra MARTINS	CR1, Université de Rouen Normandie	Directrice de thèse
Mme Pascaline GAILDRAT	MCF, Université de Rouen Normandie	Co-encadrante de thèse
M. Thierry FREBOURG	PU-PH, Université de Rouen Normandie	A titre honorifique

**Thèse dirigée par Dr. Alexandra MARTINS
Inserm U1245 – Génomique et médecine personnalisée
du cancer et des maladies neuropsychiatriques**



Résumé

Aujourd'hui, le défi majeur en génétique médicale est l'interprétation clinique des variations détectées dans le génome des patients. Dans un contexte de maladie monogénique, l'identification de la variation pathogène permet d'optimiser la prise en charge des patients et de leurs apparentés. Les variations non-sens ainsi que celles localisées au niveau des sites canoniques d'épissage (IVS±1/2) sont généralement considérées causales. Toutefois, il est possible que certaines d'entre elles induisent des anomalies d'épissage en phase permettant potentiellement la production d'une protéine fonctionnelle. Afin de tester cette hypothèse, nous avons choisi comme modèles d'étude deux gènes majeurs de prédisposition aux cancers : *BRCA2*, pour le syndrome sein-ovaire et *MSH2*, pour le syndrome de Lynch. Des approches expérimentales complémentaires ont été mises à profit afin de caractériser l'impact de ce type de variations non seulement au niveau de l'épissage mais également au niveau de la protéine. Notre étude portant sur des variations de *BRCA2* a démontré que certaines variations IVS±1/2 et non-sens (i) induisent un saut en phase de l'exon 12 du fait de l'altération de sites d'épissage ou d'éléments de régulation et (ii) sont de ce fait responsables du maintien de l'activité *BRCA2*, tout en étant hypomorphes. Ces travaux montrent, pour la première fois dans un gène de prédisposition au cancer, que certaines variations présumées nulles peuvent contourner la perte totale de fonction du fait de leur impact sur l'épissage. Nos données remettent ainsi en question le caractère pathogène de ces variations. Dans le cadre de notre étude portant sur des variations IVS±1/2 de *MSH2*, nous avons caractérisé trois biotypes d'anomalies d'épissage en phase : (i) des sauts d'exon, (ii) des délétions de portions exoniques et (iii) des rétentions introniques, avec pour conséquence au niveau protéique des délétions ou insertions de plus ou moins grande taille dans différents domaines fonctionnels. Toutes les isoformes protéiques *MSH2* résultant de ces variations splicéogéniques se sont avérées inactives, confirmant ainsi leur caractère pathogène. L'ensemble de ces travaux souligne l'importance d'exercer une certaine prudence dans l'interprétation des variations considérées *a priori* pathogènes mais susceptibles de générer des anomalies d'épissage en phase. Dans ce contexte, la combinaison d'approches expérimentales complémentaires pour appréhender l'impact biologique au niveau de l'ARN et de la protéine est essentielle pour une interprétation fiable de ce type de variations. Ces données mettent également en exergue la problématique de l'interprétation clinique des variations hypomorphes dans les gènes de prédisposition aux cancers.

Mots clés : syndrome sein-ovaire, syndrome de Lynch, *BRCA2*, *MSH2*, variations IVS±1/2 et non-sens, anomalies d'épissage en phase, interprétation des variations splicéogéniques, analyses fonctionnelles

Abstract

Today, the major challenge in medical genetics is the clinical interpretation of nucleotide variants detected in a patient's genome. In the context of a monogenic disease, the identification of the pathogenic variant allows the optimization of the medical care of patients and their relatives. Nonsense variations as well as those located at the canonical splice sites (IVS±1/2) are generally considered pathogenic. However, it is possible that a fraction of them induce in-frame splicing anomalies that can potentially result in the production of a functional protein. To test this hypothesis, we used as model systems two major cancer-predisposition genes: *BRCA2*, implicated in hereditary breast and ovarian cancer syndrome and *MSH2*, involved in Lynch syndrome. We took advantage of complementary experimental approaches to characterize the impact of this type of variants not only in RNA splicing but also at the protein level. Our study on *BRCA2* demonstrated that a subset of IVS±1/2 and nonsense variants (*i*) induce in-frame skipping of exon 12 via the modification of splice sites or of splicing regulatory elements and (*ii*) are able to maintain BRCA2 activity though being hypomorphic. These data provide the first evidence, in a cancer-predisposition gene, that certain presumed null variants can bypass total loss of function due to their impact on splicing. Thus our findings call into question the pathogenic nature of these variants. In our study on *MSH2* IVS±1/2 variants, we characterized three biotypes of in-frame splicing defects responsible for deletions or insertions at the protein level in different MSH2 functional domains: (*i*) exon skipping, (*ii*) deletions of exonic portions and (*iii*) intron retentions. All the MSH2 protein isoforms resulting from these spliceogenic variants were found to be inactive, thus confirming their pathogenic classification. Altogether, our findings highlight the need to exercise caution in the interpretation of putative pathogenic variants susceptible to induce in-frame splicing modifications. In this context, the combination of complementary experimental approaches assessing the biological impact at the RNA and protein level is essential for a reliable interpretation of this type of variants. Furthermore, our results stress the problem of the clinical interpretation of hypomorphic variants in cancer-predisposition genes.

Key words: hereditary breast and ovarian cancer syndrome, Lynch syndrome, *BRCA2*, *MSH2*, IVS±1/2 and nonsense variants, in-frame splicing defects, interpretation of spliceogenic variants, functional assay

Remerciements

Je tiens tout d'abord à remercier Madame Béatrice Parfait et Monsieur Claude Ferec de me faire l'honneur de juger ce travail et de rapporter mon manuscrit au regard de leurs expertises.

Merci également à Madame Florence Coulet de m'avoir accueillie dans son laboratoire pour ma mobilité de deux mois ainsi que d'avoir accepté d'évaluer mes travaux de thèse.

Merci à Monsieur David Tulasne pour sa participation éclairée à mes trois comités de suivi individuel de thèse ainsi que d'avoir accepté d'examiner mes travaux de thèse.

Je souhaite également remercier les autres membres de mon comité de suivi Individuel de thèse : Monsieur Gérard Le Gac, Madame Myriam Bernaudin et Monsieur Alexandre Mutel. Ce fut un plaisir de discuter de mes travaux avec vous et je vous remercie pour vos conseils avisés.

Je remercie Madame Martine Muleris d'avoir accepté de me former à son test fonctionnel avec tant de pédagogie et d'avoir facilité ma mobilité à Paris. Merci à Madame Marie Christine Waill pour m'avoir également formé et intégré à la vie du laboratoire pour quelques semaines.

Je tiens à remercier Monsieur Thierry Frebourg, directeur de l'unité Inserm U1245, de m'avoir accueillie dans son laboratoire et son équipe en tant que stagiaire puis en tant que doctorante. Merci pour tous les bons conseils prodigués.

Un remerciement tout particulier à ma directrice de thèse Alexandra Martins et ma co-encadrante de thèse Pascaline Gaildrat pour avoir accepté de tenter cette aventure qu'est la thèse avec moi. Merci Pascaline de m'avoir encadré depuis le début quand je n'étais encore qu'en Master 1. Je vous remercie infiniment toutes les deux pour le soutien dont vous avez su faire preuve au cours des années et de la confiance que vous m'avez accordée pour mener à bien nos travaux. Cela n'aura pas toujours été facile mais je suis sortie grandie de cette épopée et c'est grâce à vous. Merci de m'avoir sortie de ma zone de confort et de m'avoir emmenée en Suède ou encore aux Assises de Génétique, parler devant tant de monde de nos résultats. Merci pour les nombreuses discussions scientifiques et tout le temps que vous m'avez consacré. Votre encadrement rigoureux et attentionné m'ont permis d'évoluer scientifiquement mais également en tant que personne. Cela m'a fait très plaisir de travailler avec vous.

Aurélie, merci pour les nombreux échanges, la gentillesse, la bonne humeur et le soutien. Cela a été très chouette de faire partie du même groupe. Bon courage avec tes SNaPshot !

Françoise, ma chère Maître Jedi de la manip, merci pour toutes ces séances de remue-méninges où on s'est creusé la tête pour élucider mes problèmes expérimentaux. Merci pour ta gentillesse et surtout, ta bienveillance. J'ai hâte de repasser faire un babyfoot à Catenay !

Anne-Claire, merci pour ta bonne humeur, ton humour, ton aide sans faille et toutes les discussions, scientifiques ou autres. Merci d'avoir toujours été positive et surtout, merci d'avoir traqué les erreurs de mon manuscrit.

Raphaël et Steeve, mes anciens voisins de bureau, merci pour la bonne humeur permanente et les discussions enflammées autour du café matinal.

Un grand merci à tous les membres de l'unité U1245 pour l'entraide et la sympathie. Je ne saurai tous les citer de peur d'en oublier. Un merci spécial tout de même pour Nathalie, pivot de cette unité, qui a toujours la réponse à toutes les questions.

Je souhaiterai également remercier tous les anciens, ceux qui ne sont plus au laboratoire mais qui ont tant comptés. Merci Omar-chou d'avoir été la personne la plus douce et la plus à l'écoute, merci d'avoir toujours été présent pour m'aider. Merci Sabine d'avoir été ma thésarde ainée par procuration, de m'avoir soutenue lors de mon premier congrès et d'être toujours là quand on en a besoin. Cher Alex, merci pour ton calme olympien, tes très bons conseils et ton humour, qu'est-ce que l'on a ri ensemble.

Marinette, merci pour ta folie et ton énergie sans limite ! Merci d'avoir apporté tant de soleil au laboratoire, il a toujours fait un peu plus gris depuis que tu es parti. J'ai adoré te croiser tôt le matin au laboratoire, perchée sur tes roulettes pour aller toujours plus vite en manip. Merci pour tous ces bons moments à la PO, merci de m'avoir fait remonter à cheval et d'être toujours présente, même de l'autre côté de l'océan.

Maud La Rouge, merci d'être qui tu es, entière, drôle, passionnée et surtout, une sacrée joueuse de babyfoot ! Même si tu n'es plus tout prêt toi aussi, tu resteras l'une des plus belles découvertes qu'il m'a été donné de faire pendant ma thèse. Merci d'avoir toujours été fun et folle et merci pour toutes les belles discussions.

Je tiens également à remercier toutes les personnes qui sont passées par le bureau que j'ai eu la chance d'occuper pendant 4 ans. François (Tchoupi), merci pour ton humour décapant et ta capacité à faire disparaître la nourriture plus vite que ton ombre. Sabrina, merci d'avoir été une si belle personne et de nous avoir toujours couvert d'amour le peu de temps où tu es passé. Morganette, merci d'avoir apporté ton humour, ta bonne humeur et des photos de ton chien. J'ai hâte que l'on se retrouve enfin autour d'un jeu de société. Pauline, merci pour le soutien infaillible, les discussions interminables et les nombreuses sorties au glacier du coin. Merci d'avoir été là quand j'en avais besoin et d'être toujours prête à sortir manger quelque part pour discuter.

Camille, ma chère binôme, merci d'avoir fait un bout de chemin avec moi et de m'avoir épaulé pendant cette aventure. Cela n'aura pas toujours été simple, mais on aura bien rigolé quand même.

Sigrid et Corentin (mes enfants terribles), merci d'avoir été deux lurons complètement zinzins qui ont égayés mes journées. Vous êtes deux personnalités brillantes et vous m'en avez mis plein les yeux.

Thomas, mon cher Bipou, je voudrais te remercier pour avoir été un si bon ami pendant tout ce temps. A l'écoute, toujours positif et si patient. Merci d'avoir été là tout au long de cette thèse et d'avoir toujours été de très bons conseils. Je sais déjà que les débats du mercredi midi me manquent, j'espère qu'on aura l'occasion de toujours les faire vivre en dehors du laboratoire. Merci d'être la personne que tu es, cela a été un véritable plaisir de travailler à tes côtés.

Et bien sûr, je souhaiterai remercier l'ensemble des personnes qui me sont proches, qui ont vécu cette aventure avec moi et qui sont des membres à part entière de cette histoire. Merci à Cyril, Marine, Mushu, Valentin, Indra et les copains de NormaDoc de m'avoir tant apporté au fil des ans.

Marie, merci pour le soutien sans bornes, les longues discussions autour d'un brunch et les bons moments de rigolades. Merci à toi de m'avoir un jour proposé de participer avec toi et Marie-Anne à une présentation pendant le Master 1, merci pour les doux moments à Deauville et pour toutes les balades dans Rouen. On aura vécu nos thèses à distance mais au même moment, et on aura réussi à y arriver ensemble. Je suis si fière que l'on ait partagé cela toutes les deux.

Graciane, Gilbert, Juliana, merci pour tous les moments de bonheur partagés et toutes les jolies intentions du quotidien qui permettent de toujours trouver la motivation.

A mes parents et ma petite sœur, je les remercie d'avoir toujours veillé sur moi, même de loin, et de m'avoir toujours encouragé. Je sais que ce travail de thèse résonnera tout particulièrement pour ma maman, c'est un peu pour toi que je l'ai fait.

Mes meilleurs amis, je ne saurai jamais vous remercier assez de vous avoir à mes côtés, j'espère vous avoir apportés la moitié de ce que vous m'avez donné. Et j'espère que notre grand projet se réalisera bientôt.

Robin et Diane, merci de m'avoir fait rêver pendant votre tour du monde, de m'avoir proclamé Tata (je suis si fière) et de me couvrir de brunch les dimanches matin de travail intensif.

Merci à toi mon cher Papinou d'avoir expertisé mes travaux en anglais, de m'avoir appris à bien prononcer « protein », d'être mon fournisseur officiel de caféine et de cookies, de m'avoir permis d'être Tata pour la première fois (très fière encore une fois) et pour toutes les belles choses que l'on a vécu ensemble. Trop hâte d'organiser les Papinades avec toi un jour.

Ma meilleure amie, on en aura traversé des choses depuis la 6^{ème}. Je suis si heureuse que tu fasses toujours partie de mon histoire. Merci d'avoir toujours été présente contre vents et marées, merci d'avoir entamé avec moi un million de projets, merci pour toutes les belles choses que tu as pu me dire et merci d'avoir relu l'entièreté de mon manuscrit alors que ce n'est pas ton domaine. Je ne sais pas si tout cela aurait pu être possible sans toi.

Et enfin, Vincent, ma moitié. Merci pour tout, vraiment tout. Je suis si reconnaissante de t'avoir dans ma vie. Ces dernières années nous auront vu tellement grandir : nous pacser, fêter nos dix ans ensemble, acheter notre premier apart. Merci d'avoir été si compréhensif et conciliant, merci pour ton soutien sans faille et ton humour toujours présent. Merci d'avoir cru en moi et d'avoir toujours pensé que j'en étais capable. Merci d'être la personne formidale que tu es. Et enfin, merci pour toutes les belles années à venir.

N'oublie jamais, celui qui croit savoir n'apprend plus.

Pierre Bottero

Table des matières

Liste des abréviations	1
Liste des Figures	2
Liste des Tableaux	4
Introduction	5
I. Les prédispositions héréditaires aux cancers	6
I.1. La prédisposition génétique aux cancers du sein et de l’ovaire	6
I.1.1. Les gènes majeurs de prédisposition au syndrome sein-ovaire, <i>BRCA1</i> et <i>BRCA2</i>	6
I.1.2. Les fonctions associées aux protéines BRCA	9
I.1.2.a. Le système de réparation des cassures double brin de l’ADN par recombinaison homologue	9
I.1.2.b. Les autres fonctions portées par les protéines BRCA	14
I.1.2.c. Les domaines fonctionnels des protéines BRCA	16
I.1.3. Les gènes de prédisposition au syndrome sein-ovaire non- <i>BRCA</i>	17
I.1.4. Le diagnostic du syndrome sein-ovaire	18
I.1.4.a. Identification des cas évocateurs	19
I.1.4.b. Le diagnostic moléculaire du syndrome sein-ovaire	20
I.2. Le syndrome de Lynch, prédisposition génétique au cancer colorectal	24
I.2.1. Les gènes <i>MMR</i> , gènes majeurs de prédisposition au syndrome de Lynch	25
I.2.2. Les fonctions associées aux protéines MMR	29
I.2.2.a. La réparation post-réplivative des mésappariements de l’ADN	29
I.2.2.b. La signalisation apoptotique de certains dommages de l’ADN	32
I.2.2.c. Les autres fonctions portées par les protéines MMR	35
I.2.2.d. Les domaines fonctionnels des protéines MMR	35
I.2.3. Le diagnostic du syndrome de Lynch	36
I.2.3.a. Identification des cas évocateurs	36
I.2.3.b. Le diagnostic moléculaire du syndrome de Lynch	40
II. La problématique de la classification des variations génétiques	45
II.1. Les recommandations générales dans un contexte de maladies mendéliennes	48
II.2. Les recommandations spécifiques dans le contexte du syndrome sein-ovaire	58
II.3. Les recommandations spécifiques dans le contexte du syndrome de Lynch	64
III. L’épissage, étape clé de la maturation des ARN pré-messagers	70
III.1. La réaction catalytique d’épissage	71
III.2. La machinerie d’épissage : le splicéosome	72
III.2.1. Le splicéosome majeur	73
III.2.1. Le splicéosome mineur	76
III.3. Les signaux d’épissage	76
III.3.1. Les signaux principaux	76
III.3.2. Les signaux auxiliaires	79
III.4. L’épissage alternatif de l’ARNm	80
III.4.1. La régulation de l’épissage alternatif	82
III.4.2. Le rôle de l’épissage alternatif dans la régulation de l’expression génique	83
III.4.3. Le système de surveillance des ARNm : <i>nonsense mediated decay</i> ou NMD	84
III.4.4. L’épissage alternatif des gènes <i>BRCA</i> et <i>MMR</i>	85

IV. Les défauts d'épissage induits par des variations génétiques	91
IV.1. Les mécanismes d'altération de l'épissage	93
IV.1.1. Altération des signaux <i>cis</i> d'épissage.....	94
IV.1.1.a. Altération des signaux principaux.....	95
IV.1.1.b. Altération des signaux auxiliaires	98
IV.1.2. Altération des facteurs <i>trans</i> d'épissage	100
IV.2. Les outils de prédiction de l'impact des variations splicéogéniques.....	101
IV.2.1. Outils de prédiction dédiés aux sites d'épissage	102
IV.2.2. Outils de prédiction dédiés au site de branchement.....	105
IV.2.3. Outils de prédiction dédiés aux éléments de régulation de l'épissage	107
IV.2.4. Outils de prédiction combinés	110
IV.3. Les approches expérimentales de détection des anomalies d'épissage	111
IV.3.1. Analyses expérimentales à partir du matériel biologique des patients.....	111
IV.3.1.a. Analyse des profils d'épissage des transcrits par RT-PCR	113
IV.3.1.b. Analyse de l'expression allélique	115
IV.3.1.c. Analyses des défauts d'épissage par la technique de RNA-seq	116
IV.3.2. Analyses expérimentales alternatives	117
IV.3.2.a. Test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes	117
IV.3.2.b. Test indicateur d'anomalies d'épissage basé sur l'utilisation de transgènes ou de modifications du génome	120
Objectifs des travaux de thèse	124
Résultats	127
I. Toutes les variations non-sens ne conduisent pas à une perte totale de fonction : le paradigme de l'exon 12 de <i>BRCA2</i>	128
II. Caractérisation fonctionnelle des variations à l'origine d'une anomalie en phase de l'épissage : le modèle <i>MSH2</i>	163
Discussion	201
I. Les variations présumées pathogènes sont susceptibles de maintenir une fonction par le biais de leur impact sur l'épissage	203
I.1. Les variations IVS±1/2 responsables d'un défaut d'épissage en phase et leurs conséquences fonctionnelles au niveau de la protéine.....	203
I.1.1. Les variations IVS±1/2 à l'origine d'un saut d'exon en phase	203
I.1.2. Les variations IVS±1/2 à l'origine de délétions exoniques ou de rétentions introniques en phase.....	206
I.1.3. Autres mécanismes à l'origine du maintien d'une fonction dans le contexte de variations IVS±1/2.....	207
I.2. Les variations non-sens à l'origine d'un épissage en phase et responsables du maintien d'une fonction au niveau de la protéine.....	208
I.2.1. Les variations non-sens à l'origine d'un saut d'exon en phase	208
I.2.2. Autres mécanismes à l'origine d'un maintien d'une fonction dans le contexte de variations non-sens.....	212
II. La problématique de l'interprétation clinique des variations hypomorphes	214
Références bibliographiques.....	218

Liste des abréviations

ADN : *Acide DésoxyRibonucléique*

ADNc : ADN complémentaire

ADNg : ADN génomique

ARN : *Acide RiboNucléique*

ARNm : ARN messenger

Pré-ARNm : ARNm précurseur

BER : *Base Excision Repair*

BP : *Branch Point*

BS : *Branch Site*

CCR : Cancer ColoRectal

DSB : *Double Strand Break*

ESE : *Exonic Splicing Enhancer*

ESR : *Exonic Splicing Regulatory elements*

ESS : *Exonic Splicing Silencer*

HBOC : *Hereditary Breast and Ovarian Cancer*

HNPCC : *Hereditary Non Polyposis Colorectal Cancer*

hnRNP : *heterogeneous nuclear RiboNucléoProtein*

ISE : *Intronic Splicing Enhancer*

ISR : *Intronic Splicing Regulatory elements*

ISS : *Intronic Splicing Silencer*

IVS : *InterVening Sequence*

mESC : *mouse Embryonic Stem Cells*

MMR : *MisMatch Repair*

MSI : *MicroSatellite Instability*

MSS : *MicroSatellite Stable*

NGS : *Next-Generation Sequencing*

NHEJ : *Non Homologous End Joining*

NMD : *Nonsense Mediated Decay*

PAF : Polypose Adénomateuse Familiale

PPT : *PolyPyrimidine Tract*

PTC : *Premature Stop Codon*

RH : Recombinaison Homologue

RNP : *RiboNucleoProteins*

SNP : *Single Nucleotide Polymorphism*

SNV : *Single Nucleotide Variant*

snRNP : *small nuclear RiboNucleoProteins*

snRNA : *small nuclear RNA*

SR : *Serine aRginine-rich*

SRE : *Splicing Regulatory Elements*

Ss : *splice site*

VSI : Variation de signification inconnue

WT : *Wild Type*

Liste des Figures

Figure 1. Représentation schématique des gènes et des domaines fonctionnels des protéines BRCA1 et BRCA2.....	7
Figure 2. Les différentes voies de réparation des cassures double brin de l'ADN	11
Figure 3. Les principaux acteurs protéiques de la réparation des cassures double brin de l'ADN par recombinaison homologue	12
Figure 4. Mécanisme de réparation des cassures double brin par recombinaison homologue.....	13
Figure 5. Les différentes fonctions des protéines BRCA1 et BRCA2 assurant le maintien de l'intégrité du génome.....	15
Figure 6. Proportion de variations pathogènes identifiées dans le génome de patients évocateurs d'un syndrome sein-ovaire à partir d'une analyse NGS multi-gènes	17
Figure 7. Arbre généalogique d'une famille évocatrice d'un syndrome sein-ovaire	19
Figure 8. Le principe de létalité synthétique dans un contexte d'inhibiteurs de PARP	23
Figure 9. Proportion des différents types de cancers colorectaux	24
Figure 10. Représentation schématique des gènes et des domaines fonctionnels des protéines MMR	26
Figure 11. Le mécanisme de réparation post-répllicative des mésappariements de l'ADN par le système MMR chez l'homme	32
Figure 12. Les deux modèles d'induction de l'apoptose à la suite de la reconnaissance d'un mésappariement O ⁶ -meG par le système MMR	34
Figure 13. Arbre généalogique d'une famille évocatrice d'un syndrome de Lynch.....	37
Figure 14. Profil d'instabilité microsatellitaire tumorale obtenu par PCR multiplex fluorescente suivi d'une électrophorèse capillaire sur séquenceur	38
Figure 15. Analyse de l'expression tumorale des protéines MMR par immunohistochimie	39
Figure 16. Recommandations au diagnostic moléculaire d'un syndrome de Lynch.....	41
Figure 17. Recommandations proposées par le réseau français pour la surveillance d'un individu porteur d'une altération pathogène d'un des gènes <i>MMR</i>	42
Figure 18. Effet génétique des variations en fonction de leur fréquence allélique.....	52
Figure 19. La majorité des variations faux-sens identifiées dans le gène <i>BRCA2</i> sont des VSI.....	60
Figure 20. Mise en évidence de molécules hybrides ADN/ARN par la technique de microscopie électronique	71
Figure 21. Les réactions de transestérification du pré-ARNm	72
Figure 22. Modèles de reconnaissance des signaux d'épissage par le splicéosome majeur	74

Figure 23. Les étapes d'assemblage du splicéosome majeur selon le modèle de définition d'intron .	75
Figure 24. Représentation schématique des signaux d'épissage principaux.....	77
Figure 25. Les différents évènements d'épissage alternatif	81
Figure 26. Le mécanisme de dégradation des transcrits porteurs d'un PTC par le système NMD.....	85
Figure 27. Représentation schématique des principaux évènements d'épissage alternatif dans les gènes <i>BRCA1</i> et <i>BRCA2</i>	87
Figure 28. Représentation schématique des évènements d'épissage alternatif dans les gènes <i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> et <i>PMS2</i>	88
Figure 29. Les différentes conséquences des anomalies d'épissage au niveau de l'ARNm mature.....	94
Figure 30. Méthode de prédiction de l'altération des éléments régulateurs de l'épissage basé sur le calcul du score $\Delta tESR_{seq}$	108
Figure 31. Principe de l'analyse de l'expression allélique par extension d'amorce.....	115
Figure 32. Principe du test fonctionnel indicateur d'anomalies d'épissage, basé sur l'utilisation d'un minigène.....	118
Figure 33. Principe de l'essai fonctionnel basé sur l'utilisation de cellules souches embryonnaires de souris	121
Figure 34. Carte représentant chaque substitution possible dans les exons 2 à 5 et 15 à 23 de <i>BRCA1</i> avec leur fonctionnalité associée déterminée par SGE.....	123

Liste des Tableaux

Tableau 1. Risques cumulés de développer une tumeur du spectre du syndrome sein-ovaire en fonction de l'âge et du gène altéré	8
Tableau 2. Recommandations proposées par le réseau français pour la surveillance d'un individu porteur d'une altération pathogène d'un gène de prédisposition au syndrome sein-ovaire	21
Tableau 3. Risques cumulés à 70 ans de développer une tumeur du spectre du syndrome de Lynch en fonction du gène altéré	27
Tableau 4. L'hétérogénéité phénotypique du syndrome de Lynch est en partie fonction de la nature du gène <i>MMR</i> altéré.....	28
Tableau 5. Recommandations internationales de bonnes pratiques standardisées : les critères d'Amsterdam II et de Bethesda révisés.....	40
Tableau 6. Variabilité interindividuelle du génome.....	46
Tableau 7. Recommandations de l'ACMG-AMP pour la classification des variations	50
Tableau 8. Bases de données de population générale et bases de données mutationnelles spécifiques des maladies mendéliennes	51
Tableau 9. Les outils bioinformatiques de prédiction de l'impact fonctionnel des variations faux-sens les plus fréquemment utilisés	56
Tableau 10. Types et nombre de variations répertoriées dans les bases de données BRCA Share™ <i>BRCA1/BRCA2</i>	60
Tableau 11. Critères de classification des variations identifiées dans les gènes <i>BRCA</i> selon le consortium ENIGMA.....	62
Tableau 12. Types et nombre de variations répertoriées dans les bases de données UMD- <i>MLH1/MSH2</i>	65
Tableau 13. Critères de classification des variations identifiées dans les gènes <i>MMR</i> selon l'organisation InSIGHT	66
Tableau 14. Estimation du pourcentage de variations splicéogéniques par rapport à l'ensemble des variations pathogènes à l'origine de maladies mendéliennes	92
Tableau 15. Exemples de maladies mendéliennes associées à une altération d'un élément <i>cis</i> ou d'un facteur <i>trans</i> d'épissage	93
Tableau 16. Exemples de variations associées à une altération d'un facteur <i>trans</i> d'épissage et responsable d'une maladie mendélienne	100
Tableau 17. Outils bioinformatiques de prédiction de l'impact des variations splicéogéniques	102

Introduction

I. Les prédispositions héréditaires aux cancers

I.1. La prédisposition génétique aux cancers du sein et de l’ovaire : le syndrome sein-ovaire

Le cancer du sein se situe au 1^{er} rang des cancers incidents chez la femme avec plus de 2 millions de nouveaux cas dans le monde en 2018, représentant ainsi près d’un quart des cas de cancers féminins (Bray *et al.*, 2018 ; <https://gco.iarc.fr>). Il est estimé qu’environ une femme sur huit développera un cancer du sein au cours de sa vie. Cette forme de tumeur maligne constitue la première cause de mortalité féminine par cancer dans le monde en 2018, avec plus de 600 000 décès. En France, cela représente plus de 56 000 nouveaux cas et plus de 13 000 décès pour l’année 2018 (<https://gco.iarc.fr>). Le cancer de l’ovaire est 7 fois moins fréquent mais, du fait de son diagnostic tardif à un stade avancé, il représente la huitième cause de mortalité féminine par cancer avec près de 185 000 décès dans le monde en 2018 (Bray *et al.*, 2018 ; <https://gco.iarc.fr>). En France, pour l’année 2018 c’est près de 5 000 nouveaux cas qui ont été identifiés et près de 4 000 décès (<https://gco.iarc.fr>). La majorité des cas de cancers du sein et de l’ovaire sont sporadiques, c’est-à-dire qu’ils résultent d’anomalies somatiques acquises au cours de la vie de l’individu. Cependant, 5 à 10% des cancers du sein et de l’ovaire sont le résultat d’une prédisposition génétique héréditaire, ils sont alors connus sous le nom de syndrome sein-ovaire ou syndrome HBOC (*Hereditary Breast and Ovarian Cancer*) (Mahdavi *et al.*, 2018).

C’est en 1866 que Paul Broca, médecin français, décrit l’un des premiers cas de forme héréditaire de cancer du sein. En effet, il fait l’observation de l’histoire particulière de la famille de son épouse, 10 femmes diagnostiquées d’un cancer du sein sur 3 générations successives, et le commente dans son rapport « Traité des tumeurs » (Broca, 1866 ; Rahman, 2014). Environ cent ans plus tard, le groupe de Henry Lynch, médecin américain, fait état de plusieurs cas de cancer du sein et de l’ovaire s’agréant à un âge précoce dans différentes familles. Cette analyse a permis de démontrer l’existence d’une prédisposition génétique aux cancers du sein et de l’ovaire associée à un mode de transmission autosomique dominant (Lynch and Krush, 1971 ; Lynch *et al.*, 1972 ; Lynch *et al.*, 1978 ; Lynch *et al.*, 2013).

I.1.1. Les gènes majeurs de prédisposition au syndrome sein-ovaire, *BRCA1* et *BRCA2*

Une avancée majeure dans la connaissance de l’étiologie du syndrome sein-ovaire a été l’identification, par des études de liaison génétique et de clonage positionnel, des deux gènes majeurs de prédisposition, *BRCA1* (*BRCA1* (*BRCA1*)) (MIM #113705) (Miki *et al.*, 1994) et *BRCA2* (*BRCA2* (*BRCA2*))

2) (MIM #600185) (Wooster *et al.*, 1995). Le gène *BRCA1* est localisé dans la région 17q21.3 et se compose de 24 exons dont 22 d'entre eux codent une protéine nucléaire de 1 863 acides aminés (Figure 1). Le gène *BRCA2*, quant à lui, est localisé dans la région 13q13.1 et est composé de 27 exons dont 26 codent une protéine nucléaire de 3 418 acides aminés (Figure 1). Les gènes *BRCA1* et *BRCA2* présentent des caractéristiques communes, comme un premier exon non codant ou encore, un exon central codant plus de 50% de la protéine (exon 11). Cependant, ils ne présentent aucune homologie de séquence et codent des protéines aux fonctions distinctes. Bien que l'expression des gènes *BRCA1* et *BRCA2* soit ubiquitaire, les altérations constitutionnelles pathogènes de ces deux gènes augmentent fortement le risque de cancer surtout au niveau des tissus mammaires et ovariens. La raison de cette spécificité tissulaire demeure encore inexpliquée. Malgré tout, quelques hypothèses ont été avancées parmi lesquelles la présence de mécanismes de réparation redondants dans les autres tissus (Sedic *et al.*, 2015), la perte d'hétérozygotie très fréquente dans les tissus mammaires et ovariens (Clarke *et al.*, 2006), l'influence des hormones de type œstrogène (Chen, 2008) ou encore, l'impact de facteurs cancérigènes non génétiques (Néné *et al.*, 2019 ; Singh and Yu, 2020).

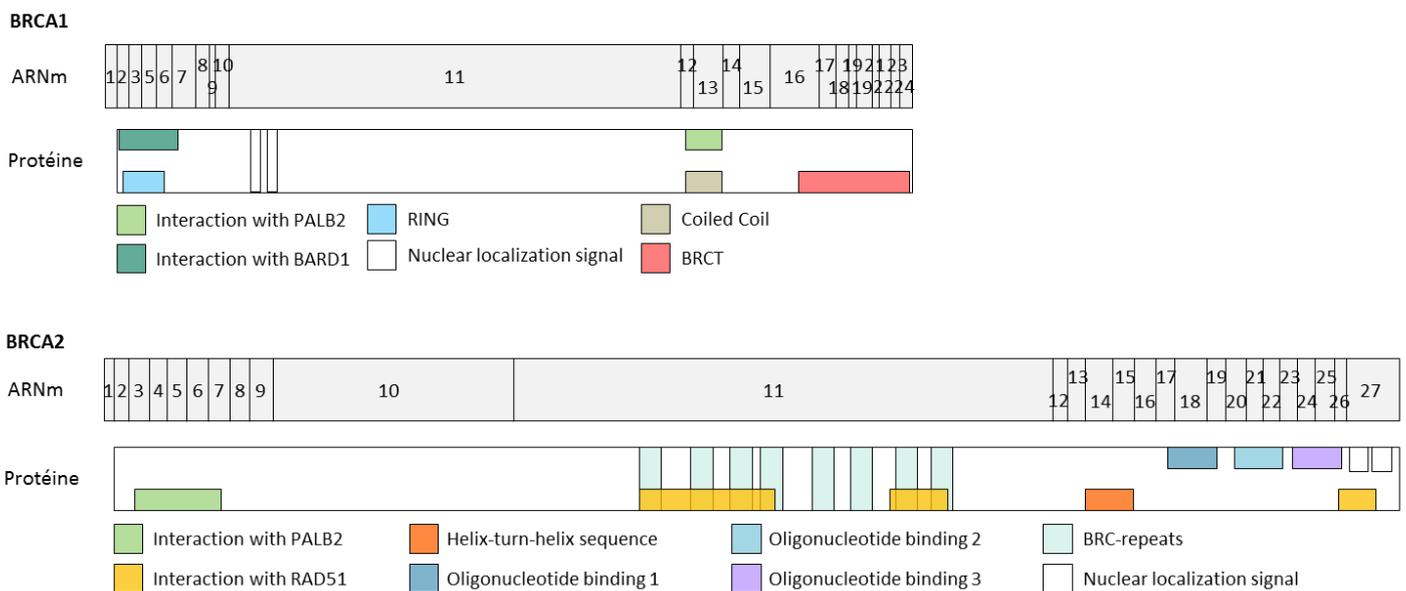


Figure 1. Représentation schématique des gènes et des domaines fonctionnels des protéines BRCA1 et BRCA2 (adapté de Liu *et al.*, 2012)

Les altérations constitutionnelles hétérozygotes pathogènes des gènes *BRCA* sont caractérisées par une forte pénétrance toutefois incomplète et variable selon le gène altéré, l'âge et le tissu considéré. Ainsi, une anomalie responsable de la perte de fonction d'un de ces gènes sera à l'origine d'un risque plus au moins important de développer une tumeur du spectre, sans induire un

risque total. Le risque cumulé d'atteinte d'un cancer du sein à l'âge de 80 ans est de 72% et 69% chez les femmes porteuses d'une variation pathogène de *BRCA1* et *BRCA2*, respectivement (Tableau 1 ; Kuchenbaecker *et al.*, 2017). Ce risque augmente très rapidement entre l'âge de 30 et 40 ans pour les porteuses d'une variation de *BRCA1*, et entre 40 et 50 ans pour les porteuses d'une variation de *BRCA2*. Egalement, une femme ayant été atteinte d'un premier cancer du sein, présentera 40% et 26% de probabilité de développer un cancer du sein controlatéral 20 ans après le premier diagnostic, si celle-ci est porteuse d'une variation pathogène de *BRCA1* ou *BRCA2*, respectivement. De plus, le risque cumulé d'atteinte d'un cancer de l'ovaire à l'âge de 80 ans est de 40% et 17% chez les femmes porteuses d'une variation pathogène de *BRCA1* et *BRCA2*, respectivement (Tableau 1 ; Kuchenbaecker *et al.*, 2017). Il est également à noter qu'une anomalie de *BRCA2* détectée chez un individu, quel qu'en soit le sexe, conduira à un risque supplémentaire d'environ 5% de développer un mélanome ou un cancer du pancréas. Tandis que chez les hommes, une anomalie de *BRCA2* conduira à un sur-risque de cancer du sein et également un risque supplémentaire d'environ 5% de développer un cancer de la prostate (Cohen-Haguenaer, 2019).

Breast Cancer		Ovarian Cancer	
<i>BRCA1</i> mutation carriers		<i>BRCA1</i> mutation carriers	
≤20		≤20	
21-30	4 (2-7)	21-30	
31-40	24 (21-29)	31-40	2 (1-3)
41-50	43 (39-48)	41-50	8 (6-12)
51-60	56 (51-61)	51-60	20 (16-26)
61-70	66 (61-72)	61-70	41 (33-50)
71-80	72 (65-79)	71-80	44 (36-53)
<i>BRCA2</i> mutation carriers		<i>BRCA2</i> mutation carriers	
≤20		≤20	
21-30	4 (2-9)	21-30	
31-40	13 (9-19)	31-40	0 (0-2)
41-50	35 (29-41)	41-50	0 (0-2) ^e
51-60	53 (46-59)	51-60	7 (4-11)
61-70	61 (55-68)	61-70	15 (10-23)
71-80	69 (61-77)	71-80	17 (11-25)

Tableau 1. Risques cumulés de développer une tumeur du spectre du syndrome sein-ovaire en fonction de l'âge et du gène altéré (adapté de Kuchenbaecker *et al.*, 2017). Les chiffres à gauche représentent un intervalle d'âge et ceux de droite indiquent le risque exprimé en pourcentage.

Les protéines codées par les gènes *BRCA1* et *BRCA2* sont impliquées dans une des voies de réparation des lésions de l'ADN et jouent un rôle important dans le maintien de l'intégrité du génome. A ce titre, les gènes *BRCA* sont considérés comme des gènes suppresseurs de tumeurs et donc, répondent au modèle « *two-hit* » de Knudson. Selon ce modèle, les deux allèles du gène doivent être inactivés pour rendre la cellule vulnérable à l'instabilité génomique et induire ainsi la tumorigenèse. Le premier *hit* correspond à l'inactivation du premier allèle, conséquence d'une altération constitutionnelle pathogène perte de fonction de l'un des gènes *BRCA* tandis que le second *hit* survient

sur le deuxième allèle dans les cellules à l'origine du tissu tumoral, sous la forme d'une altération somatique acquise au cours de la vie. Ce deuxième *hit* peut correspondre à une perte d'hétérozygotie, une variation ponctuelle pathogène, ou encore une altération épigénétique comme une hyperméthylation du promoteur (Van Heetvelde *et al.*, 2018). Du fait de l'inactivation des deux allèles, la voie de réparation dépendante de la fonction BRCA n'est plus fonctionnelle et des lésions s'accumulent dans le génome de la cellule modifiant ainsi son phénotype basal vers un phénotype tumoral.

I.1.2. Les fonctions associées aux protéines BRCA

Les protéines BRCA1 et BRCA2 jouent un rôle majeur dans la réparation des lésions de l'ADN de type cassures double brin par recombinaison homologue, mais elles sont également impliquées dans de nombreux autres processus cellulaires, tels que la régulation de la transcription, le remodelage de la chromatine ou encore, la régulation du cycle cellulaire (Roy *et al.*, 2011).

I.1.2.a. Le système de réparation des cassures double brin de l'ADN par recombinaison homologue

La réplication de l'ADN et la nécessité de maintenir l'intégrité du génome

Maintenir un faible taux de mutations dans l'ADN et assurer l'intégrité de l'information génétique est indispensable au bon développement et fonctionnement d'un organisme et dépend essentiellement de l'efficacité du processus de réplication de l'ADN (Bebeneck and Ziuzia-Graczyk, 2018). La fidélité de la réplication de l'ADN est assurée par les ADN polymérases, responsables de la synthèse du nouveau brin d'ADN à partir du brin matrice par ajout des bases nucléiques en fonction de la complémentarité des appariements (A-T et G-C). Il est estimé que les ADN polymérases introduisent une erreur toutes les 10^4 - 10^5 nucléotides polymérisés, ce qui représente un taux d'erreur très faible (Kunkel and Loeb, 1981). Ces polymérases sont capables de reconnaître spécifiquement la géométrie d'un appariement correct entre deux bases et de ralentir la réaction de réplication en cas de mauvais appariements. Ainsi, elles sont capables de corriger leurs propres erreurs grâce à leur activité exonucléase 3'-5' permettant l'élimination du dernier nucléotide incorporé et la resynthèse correcte du nouveau brin. Cette étape, appelée *proofreading*, augmente la fidélité de la synthèse de l'ADN d'un facteur de 10^2 - 10^3 , soit une erreur toutes les 10^7 nucléotides polymérisés environ (Shevelev and Hübscher, 2002). Malgré l'efficacité de réplication des ADN polymérases et leur mécanisme de contrôle, des erreurs peuvent survenir au sein de la molécule d'ADN.

Les lésions de l'ADN et les différents mécanismes de réparation

L'échappement au système de *proofreading* des ADN polymérases n'est pas la seule source de lésions de l'ADN ; celles-ci peuvent également être d'origine endogène ou exogène (Mehta and Haber, 2014). En effet, il est possible qu'au sein même de la cellule l'exposition aux espèces réactives de l'oxygène, les erreurs de méthylation, le stress mécanique ainsi que les dépurinations et dépyrimidations conduisent à des lésions de l'ADN. L'exposition à des agents génotoxiques physiques (radiations ionisantes, désamination à la suite d'un excès de chaleur) ou chimiques (espèces réactives de l'oxygène, chimiothérapies, agents alkylants) peut également être à l'origine de lésions de l'ADN. Les dommages résultants peuvent prendre différentes formes selon l'origine de la lésion ; on peut notamment citer les cassures simple brin ou double brin de l'ADN (DSB, *Double Strand Break*) (Mehta and Haber, 2014). Les DSB sont les dommages de l'ADN les plus délétères car ils touchent les deux brins et sont rétrovirogènes pour la viabilité cellulaire s'ils ne sont pas correctement corrigés. En effet, l'accumulation de ces cassures peut causer un arrêt du cycle cellulaire et par la suite la mort cellulaire.

Dans un contexte de DSB, deux voies de réparation dominent : la réparation par jonction des extrémités non-homologues (NHEJ, *Non Homologous End Joining*) ou la réparation par recombinaison homologue (RH) (Figure 2 ; Ceccaldi *et al.*, 2016 ; Scully *et al.*, 2019). Le NHEJ dit classique (cNHEJ) est à distinguer du système alternatif de ligature des extrémités non homologues (aEJ, *alternative End Joining*) médié par la micro-homologie et fonctionnant comme le cNHEJ mais sans les facteurs protéiques associés à ce système. La réparation par cNHEJ est un mécanisme rapide permettant la ligature directe des deux extrémités de la cassure indépendamment de l'homologie de séquence ; c'est-à-dire que ce système assurera la continuité de la molécule d'ADN sans restaurer la séquence initiale. De ce fait, le cNHEJ génèrera fréquemment des erreurs dans la séquence d'ADN pouvant, *in fine*, entraîner une instabilité du génome. Le système de RH, quant à lui, va assurer l'homologie de séquence en se servant de la chromatide sœur porteuse de la séquence homologue à celle endommagée. Outre ces deux mécanismes majeurs de réparation des DSB, il existe d'autres voies alternatives comme l'aEJ et une voie permettant la liaison via un appariement de bases entre deux séquences homologues de l'ADN de chaque côté de la cassure conduisant à la perte de la séquence entre ces deux régions (SSA, *Single Strand Annealing*) (Figure 2 ; Ceccaldi *et al.*, 2016 ; Scully *et al.*, 2019). Ces deux voies conduisent, comme le cNHEJ, à l'incorporation d'erreurs dans le génome de la cellule. Ainsi, le mécanisme de RH constitue le système de réparation des cassures double brin de l'ADN le plus fidèle.

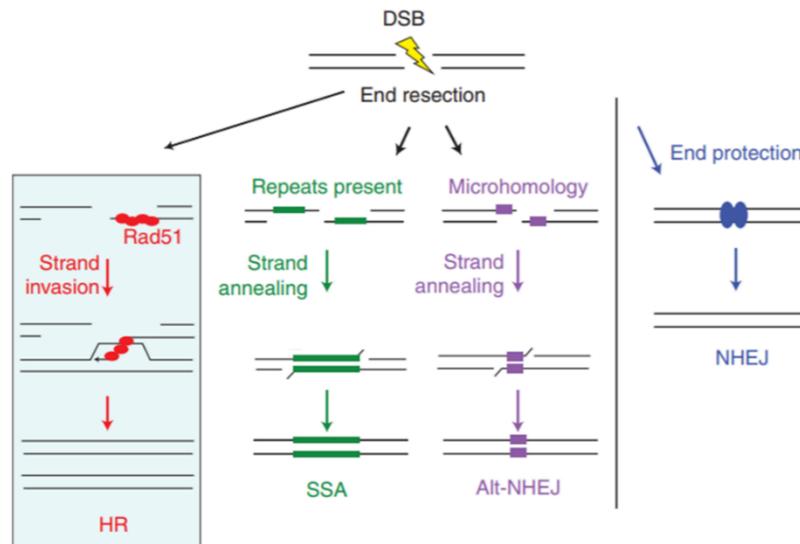


Figure 2. Les différentes voies de réparation des cassures double brin de l'ADN (d'après Jasin and Rothstein, 2013)

L'intervention d'une voie par rapport à une autre est déterminée par plusieurs facteurs dont principalement la phase du cycle cellulaire dans laquelle la lésion est détectée (Ceccaldi *et al.*, 2016 ; Scully *et al.*, 2019). L'entrée dans chaque phase du cycle cellulaire est dépendante de certaines voies de régulation constituant les points de contrôles appelés *checkpoints*. Lorsque la cellule présente une lésion de l'ADN, les *checkpoints* induisent le ralentissement ou l'arrêt de la progression du cycle cellulaire et activent en parallèle les mécanismes de réparation appropriés, afin d'empêcher le transfert des erreurs aux cellules filles (Hustedt and Durocher, 2017). Étant donné leur fonction intrinsèque, certaines voies de réparation ne sont pas actives durant toutes les phases du cycle cellulaire. En effet, si le NHEJ peut intervenir tout au long du cycle cellulaire, la RH ne peut être activée que pendant les phases S et G2, c'est-à-dire quand la chromatide sœur est présente.

Le système de réparation des cassures double brin de l'ADN par recombinaison homologue

Le mécanisme de réparation par RH des DSB fait appel à 3 types d'acteurs protéiques : (i) des « détecteurs », des protéines qui vont identifier les extrémités d'ADN cassés, (ii) des « effecteurs », des protéines qui vont réparer l'ADN et (iii) des « médiateurs », des protéines qui vont permettre le lien entre les premiers et les deuxièmes acteurs protéiques (Figure 3 ; Roy *et al.*, 2011).

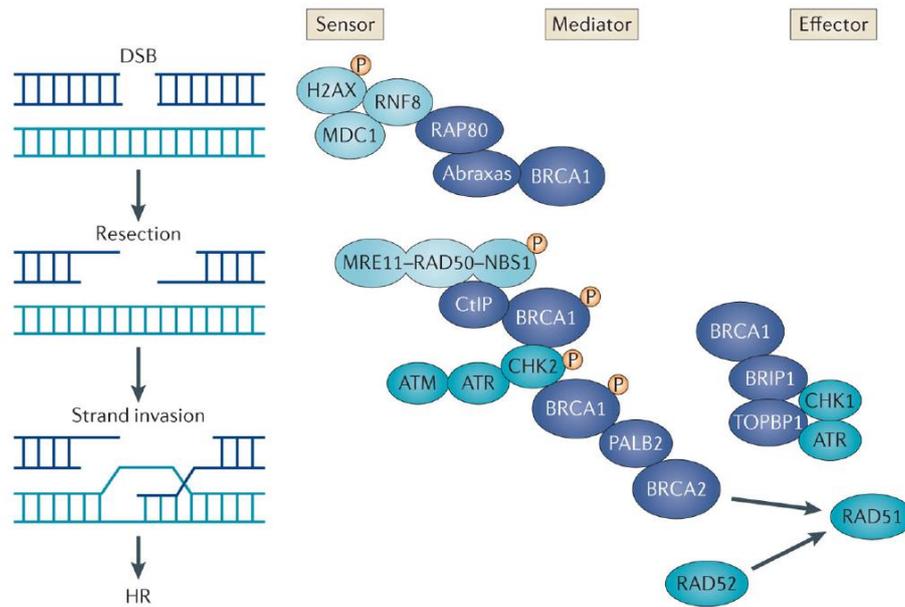


Figure 3. Les principaux acteurs protéiques de la réparation des cassures double brin de l'ADN par recombinaison homologue (d'après Roy *et al.*, 2011)

Tout d'abord, le complexe MRN formé des protéines Mre11, Rad50 et Nbs1, et la protéine CtIP, interviennent afin d'initier la résection des extrémités de la cassure (Figure 4 ; Reilly *et al.*, 2019 ; Scully *et al.*, 2019). L'activation de ce complexe induit l'activation à son tour de la kinase ATM (*Ataxia-Tealangiectasia Mutated*) qui va phosphoryler diverses protéines impliquées dans le remodelage de la chromatine, dans l'arrêt du cycle cellulaire et dans la reconnaissance des extrémités de la cassure. Les extrémités simples brins obtenues à la suite de la résection sont stabilisées par la protéine RPA (*Replication Protein A*) qui prévient également la formation de structures secondaires. Les protéines BRCA1, BRCA2, PALB2, 53BP1, RAD51 et ses homologues (RAD51B, RAD51C, RAD51D) sont ensuite recrutées afin d'amener la protéine RAD51 à former un filament nucléoprotéique à partir de l'ADN simple brin (Figure 4). Une fois celui-ci formé et l'homologie avec la chromatide sœur trouvée, l'ADN polymérase allonge l'ADN simple brin par son extrémité 3' en utilisant l'ADN complémentaire comme matrice afin de restaurer les nucléotides perdus lors de la cassure. Les extrémités seront ensuite refermées grâce à une ligase et les brins entrecroisés des deux chromatides sœurs (jonctions de Holiday) seront finalement coupés par la protéine MuS81 résultant en deux brins d'ADN intacts.

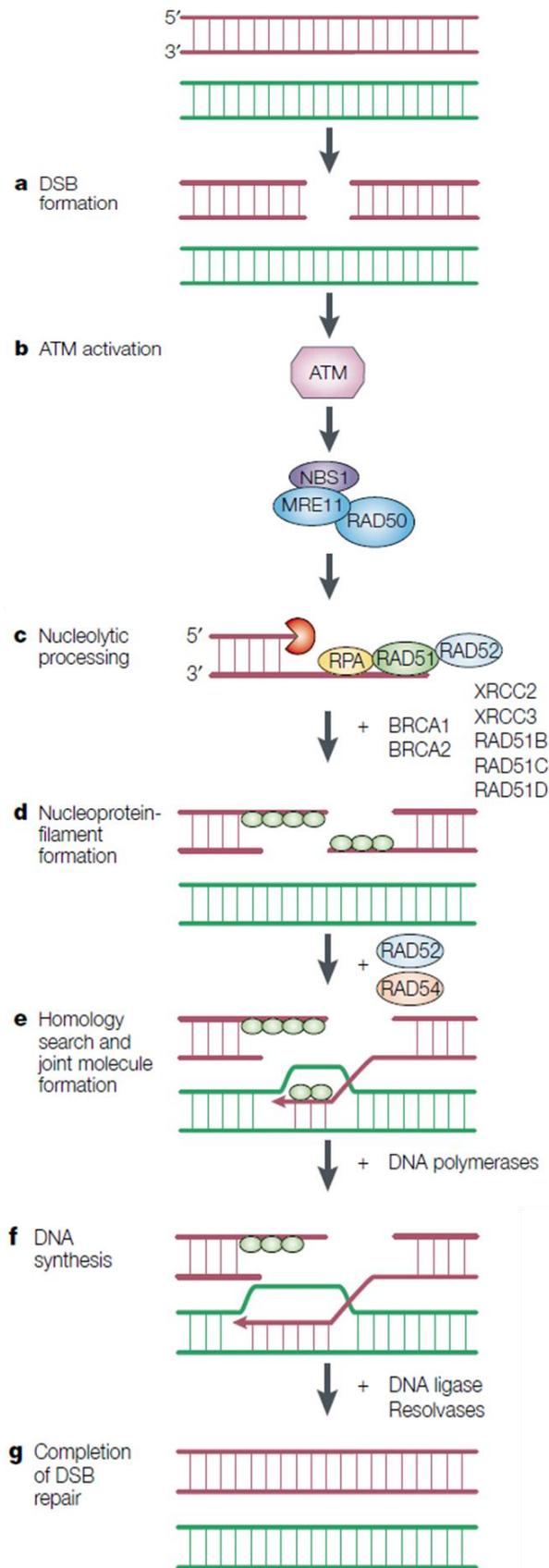


Figure 4. Mécanisme de réparation des cassures double brin par recombinaison homologue (d'après van Gent *et al.*, 2001)

Au cœur de ce mécanisme de réparation, les protéines BRCA1 et BRCA2 jouent des rôles essentiels (Figure 4 ; Gorodetska *et al.*, 2019). Tout d'abord, la protéine BRCA1 interagit directement avec la protéine CtIP phosphorylée et couplée au complexe MRN formant ainsi le complexe BRCA1-C (Prakash *et al.*, 2015 ; Chen *et al.*, 2018). Cette interaction permet de catalyser la résection des extrémités de la cassure double brin afin de générer des extrémités simple brin, étape clé dans la RH. Complémentaire à cette première fonction, la protéine BRCA1 agit également comme antagoniste de 53BP1, protéine ayant pour fonction de bloquer la résection des DSB et ainsi, par inhibition du système de RH, conduire à la mort cellulaire (Prakash *et al.*, 2015 ; Chen *et al.*, 2018). L'association de ces deux fonctions fait de la protéine BRCA1 un acteur essentiel au démarrage du mécanisme de RH. Par la suite, la protéine BRCA1 interagit avec la protéine PALB2 ce qui permet le recrutement de la protéine BRCA2 (Figure 4). Cette dernière assure un rôle majeur dans le processus de RH. En effet, la protéine BRCA2 conduit au recrutement de la protéine RAD51 sur le site de la DSB, et assure ainsi son action d'invasion des extrémités simple brin grâce (i) au déplacement de la protéine RPA qui empêche l'approche de RAD51 du fait de l'encombrement stérique et (ii) à la stabilisation de RAD51 du fait de l'inhibition de son activité ATPase ADN-dépendante (Prakash *et al.*, 2015 ; Chen *et al.*, 2018). Il a été démontré que les cellules déficientes pour BRCA1 étaient incapables de fixer PALB2, BRCA2 et RAD51 à la cassure, et que celles déficientes pour BRCA2 se montraient incapables de lier RAD51, ce qui, dans les deux cas, perturbe négativement le processus de recombinaison homologue (Prakash *et al.*, 2015). Les protéines BRCA1 et BRCA2 exercent donc des rôles différents mais indispensables au bon fonctionnement du mécanisme de recombinaison homologue.

1.1.2.b. Les autres fonctions portées par les protéines BRCA

En plus de leurs rôles essentiels dans le processus de réparation des cassures double brin de l'ADN par recombinaison homologue, les protéines BRCA sont des protéines polyvalentes assurant de nombreuses fonctions au sein de la cellule en vue de maintenir l'intégrité du génome (Figure 5 ; Martinez *et al.*, 2015 ; Gorodetska *et al.*, 2019).

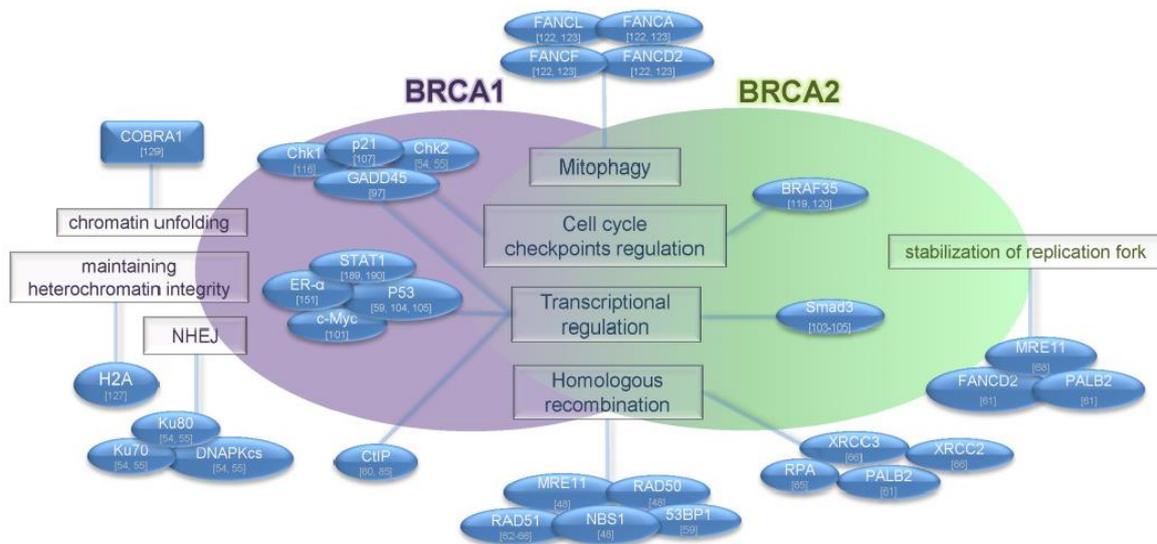


Figure 5. Les différentes fonctions des protéines BRCA1 et BRCA2 assurant le maintien de l'intégrité du génome (d'après Gorodestka *et al.*, 2019)

La protéine BRCA1 exerce la plupart de ces fonctions par l'intermédiaire d'interaction avec un très grand nombre de protéines (Figure 5). En fonction de la nature de ses partenaires protéiques, BRCA1 est capable de jouer un rôle dans différents processus cellulaires tels que (i) la régulation du cycle cellulaire, en assurant le passage d'une phase à l'autre ou l'arrêt de la progression du cycle (Ruffner and Verma, 1997 ; Wang *et al.*, 2000), (ii) la régulation de la transcription, en modulant l'activité de différents facteurs de transcription (Monteiro *et al.*, 1996 ; Nadeau *et al.*, 2000) ou encore (iii) le remodelage de la chromatine (Zhu *et al.*, 2011). On peut notamment citer son interaction avec la protéine BARD1 (*BRCA1 Associated Ring Domain 1*) qui, par formation d'un complexe hétérodimérique avec BRCA1, permet d'induire l'activité ubiquitine ligase E3 de BRCA1, à l'origine de l'ubiquitination des histones mais également de nombreuses protéines interagissant avec BRCA1 ainsi que BRCA1 elle-même (Mallery *et al.*, 2002 ; Xia *et al.*, 2003). Selon le contexte cellulaire et les signaux perçus par la protéine, l'ubiquitination induite par ce complexe peut réguler différents processus cellulaires en modulant l'activité, la dégradation ou la localisation intracellulaire de certaines protéines (Mansour, 2018). La protéine BRCA1 est également capable de réguler le stress oxydatif induit par exposition aux espèces réactives de l'oxygène (ROS, *Reactive Oxygen Stress*). Le mécanisme de cette régulation n'est que peu compris mais il a été rapporté qu'en absence de BRCA1, le niveau des ROS augmente tandis qu'une surexpression de BRCA1 fait diminuer la concentration des ROS (Saha *et al.*, 2009 ; Martinez-Outschoorn *et al.*, 2012). La protéine BRCA1 joue également un rôle essentiel dans la mitophagie, une voie spécifique d'autophagie visant à dégrader les mitochondries défectueuses. En effet, il a été démontré que l'absence de protéine BRCA1 fonctionnelle conduit à l'inhibition de la voie de mitophagie compromettant ainsi tout le réseau mitochondrial et de ce fait, la survie cellulaire (Arun

et al., 2015). La protéine BRCA1 intervient donc dans de très nombreux processus cellulaires, qu'ils soient nucléaires ou cytoplasmiques, grâce à son interaction avec une pléthore de partenaires et constitue de ce fait un gardien de l'intégrité du génome.

La protéine BRCA2, à l'instar de BRCA1, est impliquée dans de nombreux processus cellulaires tels que (i) la régulation du cycle cellulaire, en contrôlant le passage de la phase G2 à la phase M (Menzel *et al.*, 2011), (ii) la régulation de la transcription, en médiant l'activité de facteurs de transcription (Milner *et al.*, 1997) ainsi que (iii) la mitophagie (Arun *et al.*, 2015) (Figure 5). De plus, BRCA2 permet de maintenir l'intégrité des télomères ce qui prévient la dégradation des chromosomes et leur fusion. En effet, la protéine BRCA2 en favorisant le recrutement de RAD51 sur l'ADN des télomères et ainsi l'activité de RH, facilite la réplication et la protection de ceux-ci (Min *et al.*, 2012). De façon intéressante, il a été montré que dans les tumeurs déficientes en protéines BRCA2, mais pas dans celles déficientes en protéines BRCA1, les télomères sont plus courts et fragiles suggérant ainsi que l'absence de protéines BRCA2 promeut la tumorigenèse (Badie *et al.*, 2010). Enfin, la protéine BRCA2 exerce également un rôle dans la protection des fourches de réplication (Lomonosov *et al.*, 2003) et dans l'élimination des structures secondaires (Bathia *et al.*, 2014). L'ensemble de ces fonctions supporte le fait que la protéine BRCA2 est essentielle au maintien de la stabilité du génome.

1.1.2.c. Les domaines fonctionnels des protéines BRCA

Comme illustré dans la figure 1, les protéines BRCA présentent des domaines fonctionnels spécifiques, importants pour leurs différentes fonctions (Roy *et al.*, 2011). La protéine BRCA1 porte, dans sa région N-terminale, un domaine catalytique RING riche en cystéine qui permet l'interaction de BRCA1 avec la protéine BARD1 (Bienstock *et al.*, 1996 ; Brzovic *et al.*, 2001). Il a été montré que les altérations de ce domaine sont responsables de la perte d'activité ubiquitine ligase de BRCA1 et de ce fait, conduisent au développement tumoral (Hashizume *et al.*, 2001). Cette observation souligne le rôle important du domaine RING dans l'activité suppresseur de tumeur de la protéine BRCA1. Au niveau de la région C-terminale, deux domaines BRCT (*BRCA1 C-Terminal domain*) permettent l'interaction de la protéine BRCA1 avec de nombreux partenaires (Huyton *et al.*, 2000 ; Yu *et al.*, 2003). La protéine BRCA1 possède deux signaux de localisation nucléaire (NLS, *Nuclear Localization Signals*) et un domaine superhélice (*coiled coil*). Ce dernier permet à BRCA1 d'interagir avec la protéine PALB2 lors du processus de réparation des DSB par RH. La protéine BRCA2 présente deux domaines de liaison au niveau de son extrémité N-terminale : le premier permet l'interaction de BRCA2 avec la protéine PALB2, assurant ainsi le recrutement de BRCA2 dans le processus de recombinaison homologue (Xia *et al.*, 2006 ; Oliver *et al.*, 2009) tandis que le second est un domaine de liaison à l'ADN assurant

l'interaction de BRCA2 avec les cassures double brin de l'ADN (von Nicolai *et al.*, 2016). La protéine BRCA2 peut également moduler indirectement la transcription via son interaction avec le facteur de co-activation de la transcription P/CAF (*P300/CBP-Associated Factor*) (Fuks *et al.*, 1998). Localisé dans la région C-terminale, le deuxième domaine de liaison à l'ADN de BRCA2 est constitué d'un motif hélice-coude-hélice (HTH, *Helix-Turn-Helix*) et de 3 domaines OB (*Oligonucleotide Binding*) (Yang *et al.*, 2002). Le motif HTH et le premier domaine OB forment également un domaine de liaison à la protéine DSS1 (*Deleted in Split hand/Split foot 1*) conduisant à la stabilité de BRCA2 et permettant de lier le complexe protéique BRCA2-RAD51 à la protéine RPA lors du processus de recombinaison homologue (Zhao *et al.*, 2015). Dans la région C-terminale, BRCA2 possède également deux NLS et un domaine de liaison à RAD51. Au niveau de la région centrale de la protéine, 8 répétitions en tandem appelées *BRC repeats* permettent également l'interaction de BRCA2 avec RAD51, étape clé de l'activité de recombinaison homologue (Wong *et al.*, 1997 ; Carreira *et al.*, 2009). Il est toutefois important de noter que de larges régions dans les deux protéines ne sont *a priori* associées à aucune fonction, selon nos connaissances actuelles.

I.1.3. Les gènes de prédisposition au syndrome sein-ovaire non-BRCA

Les altérations des gènes *BRCA* ne peuvent expliquer à elles seules l'ensemble des cas suspectés de syndrome sein-ovaire (Kast *et al.*, 2016 ; Castéra *et al.*, 2018). En effet, il a été rapporté que seulement 24 à 51% des cas de syndrome sein-ovaire sont liés à une altération constitutionnelle d'un gène *BRCA*. Cette hérédité manquante a pu s'expliquer en partie à la suite de l'identification de variations pathogènes dans d'autres gènes qui se sont avérés impliqués dans ce syndrome (Figure 6 ; Mahdavi *et al.*, 2018 ; Angeli *et al.*, 2020).

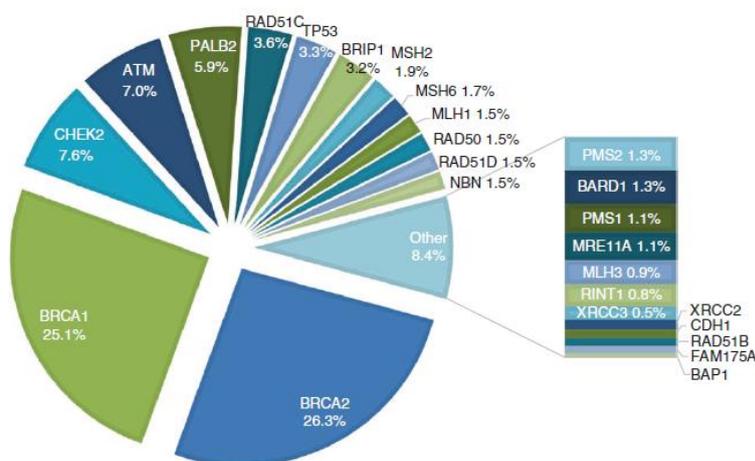


Figure 6. Proportion de variations pathogènes identifiées dans le génome de patients évocateurs d'un syndrome sein-ovaire à partir d'une analyse NGS multi-gènes (d'après Castéra *et al.*, 2018)

Plusieurs autres gènes de susceptibilité au cancer du sein ont été identifiés et associés à une forte pénétrance parmi lesquels : (i) le gène *PALB2* (*Partner And Localizer of BRCA2*) dont la fonction

protéique dans la voie de réparation des cassures double brin par recombinaison homologue est primordiale et dont les altérations constitutionnelles confèrent un risque cumulé de 44% à 70 ans (Antoniou *et al.*, 2014 ; Yang *et al.*, 2019), (ii) le gène *TP53* (*Tumor Protein 53*) impliqué dans le syndrome de Li-Fraumeni et prédisposant aux formes très précoces de cancers, avant 31 ans (Ruijs *et al.*, 2009 ; Masciari *et al.*, 2012), (iii) le gène *PTEN* (*Phosphatase and TENsin homolog*), gène suppresseur de tumeur dont l'altération conduit au syndrome de Cowden et confère un risque de 77% à 70 ans (Bubien *et al.*, 2013 ; Ngeow *et al.*, 2015) et (iv) le gène *CDH1* (*CaDHerin 1*) dont l'altération prédispose au syndrome gastrique héréditaire diffus et conduit à un risque cumulé de 40% à 70 ans (Hansford *et al.*, 2015). D'autres gènes sont plutôt associés à un risque élevé de cancer de l'ovaire dont les deux gènes paralogues de *RAD51*, *RAD51C* et *RAD51D* (Meindl *et al.*, 2010 ; Loveday *et al.*, 2016). La protéine *RAD51* exerce un rôle essentiel dans la voie de réparation des cassures double brin par recombinaison homologue et, même si ses deux paralogues n'assurent pas le même rôle, *RAD51C* et *RAD51D* régulent positivement ce mécanisme. Une altération constitutionnelle de *RAD51C* ou *RAD51D* conduit donc à une augmentation du risque cumulé de cancer de l'ovaire à 80 ans de 9% et 10%, respectivement (Sopik *et al.*, 2014 ; Loveday *et al.*, 2016). Il a également été rapporté plusieurs autres gènes de susceptibilité associés à un risque modéré comme *CHEK2*, *ATM*, *BRIP1*, *NF1* (Desrichard *et al.*, 2011 ; Tavgigian *et al.*, 2009 ; Rafnar *et al.*, 2011 ; Madanikia *et al.*, 2012). Certains gènes semblent également impliqués dans le déterminisme génétique du syndrome sein-ovaire, tels que *BARD1*, *NBN*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, mais la pénétrance reste à préciser (Easton *et al.*, 2015 ; Cohen-Haguenauer, 2019).

Le développement des technologies de séquençage à haut débit de l'ADN a permis de mieux appréhender le rôle de certains gènes non-*BRCA* dans l'étiologie du syndrome sein-ovaire. Malgré tout, il existe toujours une hérédité manquante. Celle-ci pourrait éventuellement s'expliquer par la présence d'autres gènes de prédisposition au syndrome sein-ovaire encore non identifiés ou par l'existence de variations dans des gènes de susceptibilité connus mais localisés dans des régions non séquencées.

I.1.4. Le diagnostic du syndrome sein-ovaire

Le syndrome sein-ovaire est désormais très largement documenté dans la littérature et la caractérisation de ce syndrome repose aujourd'hui sur différents éléments évocateurs tels que le contexte clinique individuel, les antécédents familiaux et les caractéristiques tumorales. Néanmoins, le diagnostic définitif de ce syndrome dépend de l'identification, dans le génome du patient, de l'anomalie constitutionnelle hétérozygote à l'origine de la perte de fonction de l'un des gènes de prédisposition.

1.1.4.a. Identification des cas évocateurs

Une prédisposition héréditaire au cancer peut être suspectée devant certains critères spécifiques. Tout d'abord, il peut s'agir d'éléments propres à l'individu tels que : (i) la précocité de survenue d'un cancer du sein isolé (avant 36 ans) ou de l'ovaire (avant 70 ans), (ii) un cancer du sein triple négatif ou médullaire avant 51 ans, (iii) la bilatéralité des cancers du sein, (iv) la multiplicité de tumeurs primitives (plusieurs cancers dans le même tissu, sein ou ovaire, ou deuxième tumeur dans un tissu différent de celui d'origine, sein, ovaire, pancréas, prostate) et (v) le sexe de l'individu (homme atteint d'un cancer du sein quel que soit l'âge) (Eisinger *et al.*, 2004 ; Cohen-Haguenaer, 2019). Le contexte familial de l'individu représente également un argument dans l'orientation du patient vers une consultation d'oncogénétique (Figure 7 ; Lynch *et al.*, 2013). Une prédisposition héréditaire est ainsi envisagée devant l'agrégation de cas de cancers du sein et/ou de l'ovaire dans la même branche parentale au premier ou au second degré avec (i) au moins 3 cas de cancers du sein et /ou de l'ovaire ou (ii) 2 cas de cancers du sein ou de l'ovaire et au moins un cas de cancer du pancréas ou (iii) 2 cas de cancers passant par un homme et ayant développés soit un cancer du sein avant 40 ou un cancer bilatéral, soit un cancer du sein et un cancer de l'ovaire, soit deux cancers de l'ovaire, soit un cancer du sein identifié chez un homme (Eisinger *et al.*, 2004 ; Cohen-Haguenaer, 2019).

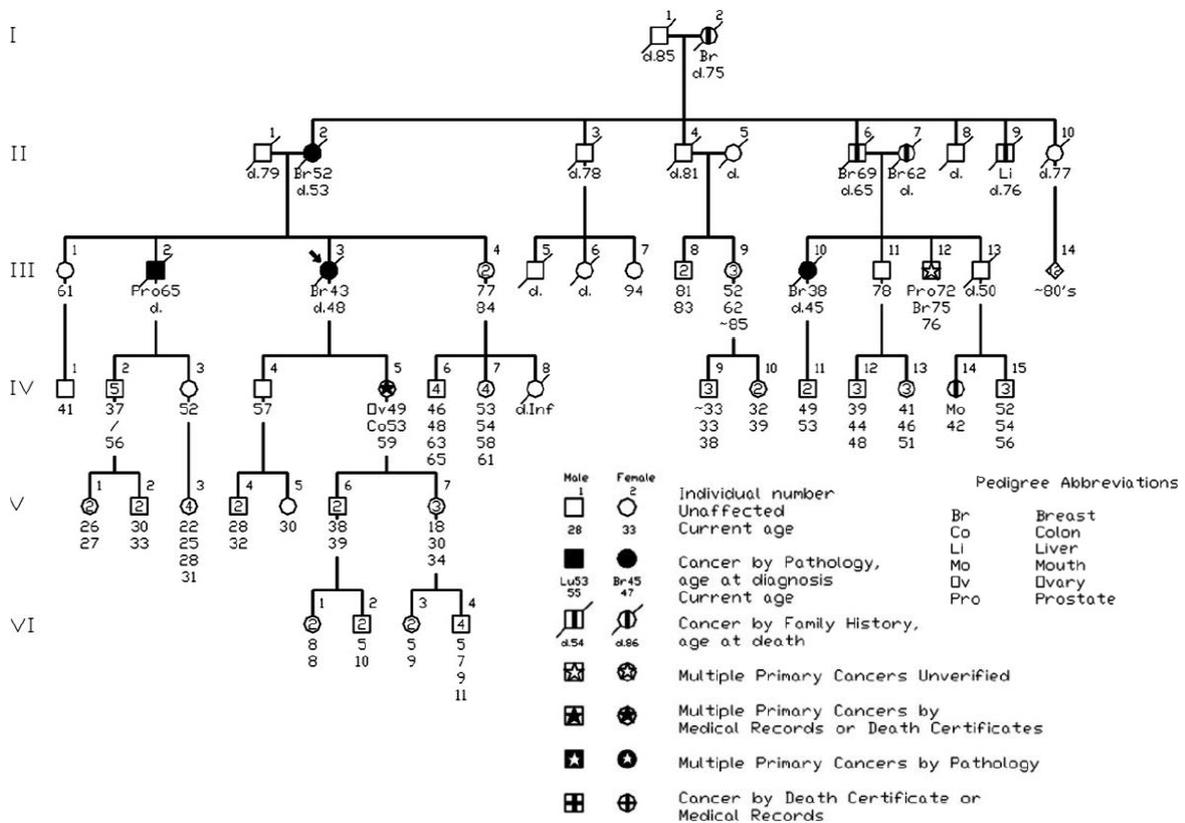


Figure 7. Arbre généalogique d'une famille évocatrice d'un syndrome sein-ovaire (d'après Lynch *et al.*, 2013)

Les cancers associés à un syndrome sein-ovaire présentent très peu de caractéristiques tumorales spécifiques par rapport aux cancers d'origine sporadique. Il est donc difficile de se baser sur les données tumorales dans le but de conclure à une prédisposition héréditaire. On peut toutefois distinguer quelques sous-types tumoraux plus fortement représentés au sein des patients présentant une forme héréditaire : le cancer du sein de type médullaire avant 50 ans, les cancers du sein de type canalaire infiltrant de grade élevé (grade III) (Chappuis *et al.*, 2000) ou les cancers dits triples négatifs (*i.e.* ayant perdu l'expression des récepteurs aux œstrogènes (RO-) et à la progestérone (RP-) et ne présentant pas de surexpression du facteur de croissance HER2 (HER2-)) (Phuah *et al.*, 2012 ; Spurdle *et al.*, 2014).

1.1.4.b. Le diagnostic moléculaire du syndrome sein-ovaire

Face à un contexte personnel et familial évocateur d'un syndrome sein-ovaire, associé ou non à l'observation d'un élément tumoral potentiellement concordant, il peut être proposé au patient un diagnostic moléculaire et ce afin de confirmer un cas de syndrome sein-ovaire. L'objectif est alors l'identification de l'anomalie constitutionnelle hétérozygote pathogène à l'origine de la perte de fonction de l'un des gènes majeurs de prédisposition. Les variations constitutionnelles bi-alléliques de *BRCA1*, *BRCA2* ou *PALB2* à l'état homozygote ou hétérozygote composite sont à l'origine d'un autre syndrome héréditaire, l'anémie de Fanconi, celui-ci à transmission autosomique récessive et caractérisé par des malformations congénitales variables, des aplasies médullaires ainsi que des hémopathies et des tumeurs solides (D'Andrea and Grompe, 2003 ; Xia *et al.*, 2007). L'identification de la variation causale est indispensable à la bonne orientation du conseil génétique pour le patient et ses apparentés, d'un point de vue diagnostique et pronostique mais également thérapeutique.

Initialement, la recherche d'une variation ponctuelle dans les régions codantes et introniques flanquantes des gènes de prédisposition s'effectuait par séquençage Sanger alors que la recherche de réarrangements génomiques de grande taille était réalisée par la technique de MLPA (*Multiplex Ligand dependent-Probe Amplification*) ou de QMPSF (*Quantitative Multiplex PCR of Short Fluorescent fragments*). Grâce à l'avènement du séquençage à haut débit (NGS, *Next-Generation Sequencing*), la recherche de l'anomalie pathogène repose désormais sur une analyse en panel ciblant simultanément plusieurs gènes (Castéra *et al.*, 2018 ; Moretta *et al.*, 2018). Cette technique a permis de révolutionner le diagnostic moléculaire du syndrome sein-ovaire en séquençant très rapidement plusieurs gènes en même temps, à moindre coût, améliorant ainsi la détection des variations (Castéra *et al.*, 2014). En France, dans un contexte de suspicion de syndrome sein-ovaire, l'anomalie pathogène sera recherchée dans les principaux gènes de prédisposition, *BRCA1* et *BRCA2*, gènes associés à un risque élevé de

cancer, mais également au sein de 11 autres gènes associés à des risques plus ou moins élevés : *PALB2*, *TP53*, *CDH1*, *PTEN*, *RAD51C*, *RAD51D*, *MLH1*, *MSH2*, *MSH6*, *PMS2* et *EPCAM*. Ce panel de 13 gènes a été établi sous l'égide du Groupe Génétique et Cancer ayant statué sur leur relevance clinique (Moretta *et al.*, 2018).

Importance du diagnostic moléculaire du syndrome sein-ovaire

L'identification de l'altération génétique à l'origine de la maladie revêt un caractère essentiel d'un point de vue diagnostique, mais également pronostique. En effet, l'identification de cette anomalie va conditionner la prise en charge du patient (cas index) mais également le suivi médical de ses apparentés. Les cliniciens pourront ainsi proposer des examens de surveillance aux individus porteurs de la variation pathogène afin de dépister précocement tout développement tumoral. Le diagnostic moléculaire chez les apparentés permet également de lever une surveillance inutile chez les individus non porteurs présentant un risque de développer une tumeur similaire à celui retrouvé dans la population générale. Des recommandations de prise en charge ont été émises par des consortiums d'experts internationaux américains et européens (Paluch-Shimon *et al.*, 2006). En France, un réseau national d'experts a été constitué afin d'établir des recommandations basées sur les données de la littérature scientifique ainsi que sur les différentes recommandations émises à l'international (Tableau 2 ; Moretta *et al.*, 2018 ; Livon *et al.*, 2019).

	Surveillance mammaire	Chirurgie de réduction des risques (CRR)		Surveillance gynécologique
		Mammaire	Pelvienne	
BRCA1 BRCA2	Sur un rythme annuel : De 30 à 65 ans IRM + Mammographie (+/- échographie). Après 65 ans Mammographie +/- Echographie (Recommandations HAS 2014 * et INCa 2017**)	Mastectomie prophylactique recevable (Recommandations HAS 2014 * et INCa 2017**)	Annexectomie prophylactique (âge à discuter à partir de 40 ans et selon la mutation et l'HF de CO) **	Avant CRR: surveillance standard, pas de dépistage ovarien efficace à préconiser **
PALB2			Pas de prise en charge gynécologique spécifique. Si HF de CO, prise en charge discutée en RCP	Surveillance standard
CDH1				Surveillance standard. Si atteinte gynécologique MC : prise en charge discutée en RCP
PTEN			Surveillance standard	
TP53	Sur un rythme annuel dès 20 ans IRM + Echographie (pas de mammographie systématique)			
RAD51C RAD51D	Pas de surveillance mammaire spécifique, à adapter à l'HF de CS selon les recommandations HAS 2014*	Non indiquée	Annexectomie prophylactique (âge à discuter à partir de 45 ans et selon l'HF de CO)	Avant CRR: surveillance standard, pas de dépistage ovarien efficace à préconiser **
MLH1 MSH2 - EPCAM MSH6 PMS2			CRR ovarienne et/ou utérine à discuter en RCP selon référentiel syndrome de Lynch	Surveillance utérine selon référentiel syndrome de Lynch

Tableau 2. Recommandations proposées par le réseau français pour la surveillance d'un individu porteur d'une altération pathogène d'un gène de prédisposition au syndrome sein-ovaire (d'après Moretta *et al.*, 2018)

Ainsi, les femmes à risque se verront proposer un suivi annuel avec un examen clinique mammaire et pelvien à partir de 30 ans. Ce suivi pourra être complété par une imagerie par résonance magnétique et une mammographie. De plus, une chirurgie prophylactique peut être proposée à ces femmes soit sous la forme d'une mastectomie bilatérale permettant de réduire le risque de cancer du sein entre 90 et 100%, soit sous la forme d'une salpingo-oophorectomie bilatérale réduisant le risque de cancer de l'ovaire entre 72 et 88% (Ludwig *et al.*, 2016). Concernant les hommes à risque, il est recommandé d'effectuer un suivi annuel avec un examen clinique mammaire et un examen prostatique.

Intérêt thérapeutique du diagnostic moléculaire du syndrome sein-ovaire

Pour les individus atteints d'un cancer, l'identification de l'altération constitutionnelle revêt un caractère essentiel d'un point de vue thérapeutique car cela impacte directement le choix du traitement médical. En effet, il a été montré que la perte de fonction de BRCA1 ou BRCA2 confère aux cellules tumorales une hypersensibilité à certaines chimiothérapies, notamment celles à base de platine (cisplatine, carboplatine) (Lee *et al.*, 2020). De façon importante, l'identification de l'altération constitutionnelle pathogène dans les gènes *BRCA1* ou *BRCA2* conditionne désormais l'accès à certaines thérapies ciblées innovantes comme celles basées sur les inhibiteurs de PARP (*Poly ADP-Ribose Polymerase*) et reposant sur le concept de létalité synthétique (Figure 8 ; Sonnenblick *et al.*, 2015 ; Lord and Ashworth, 2017). Les protéines PARP sont une famille d'enzymes dans laquelle PARP1 et PARP2, les deux principales isoformes, jouent un rôle clé dans la voie de réparation par excision de base (BER, *Base Excision Repair*) en réponse aux cassures simple brin de l'ADN (Durkaz *et al.*, 1980). L'inhibition de ces enzymes conduit à l'inactivation du système BER et de ce fait, à l'accumulation de cassures simple brin puis double brin dans l'ADN. Dans une cellule normale, sans altérations délétères de *BRCA1* et *BRCA2*, ces lésions seront réparées grâce au mécanisme de réparation des cassures double brin dépendant de la fonction BRCA. En revanche, dans une cellule tumorale porteuse de l'altération constitutionnelle, le système de réparation dépendant de la fonction BRCA est incapable de s'activer du fait de la déficience en protéines BRCA, conduisant ainsi à la mort cellulaire (Lee *et al.*, 2020). De façon intéressante, cette thérapie ciblée a également prouvé son efficacité dans le cadre d'anomalies constitutionnelles détectées au sein d'autres gènes, également impliqués dans la voie de réparation des cassures double brin tel que *PALB2* (Lord and Ashworth, 2017). En Europe, les patientes ayant développé un cancer de l'ovaire associé à une variation pathogène des gènes *BRCA* sont éligibles pour le traitement d'entretien aux inhibiteurs de PARP Olaparib, disposant d'une autorisation de mise sur le marché depuis 2014. Une étude de phase III a démontré que cette molécule conférait une amélioration de la survie sans progression tumorale de deux ans par rapport au placebo (Pujade-

Lauraine *et al.*, 2017). Plus récemment, deux autres molécules ayant démontré leur efficacité en tant que traitement d'entretien de certains cancers de l'ovaire ont également reçu une autorisation européenne de mise sur le marché : le Niraparib, en 2017 (Mirza *et al.*, 2016) et le Rucaparib, en 2018 (Coleman *et al.*, 2017). A ce jour, en Europe, seules deux molécules ont obtenu une autorisation de mise sur le marché pour le traitement de certains cas de cancer du sein après avoir démontré une amélioration de la survie sans progression tumorale : l'Olaparib (Robson *et al.*, 2017) et le Talazoparib (Litton *et al.*, 2018).

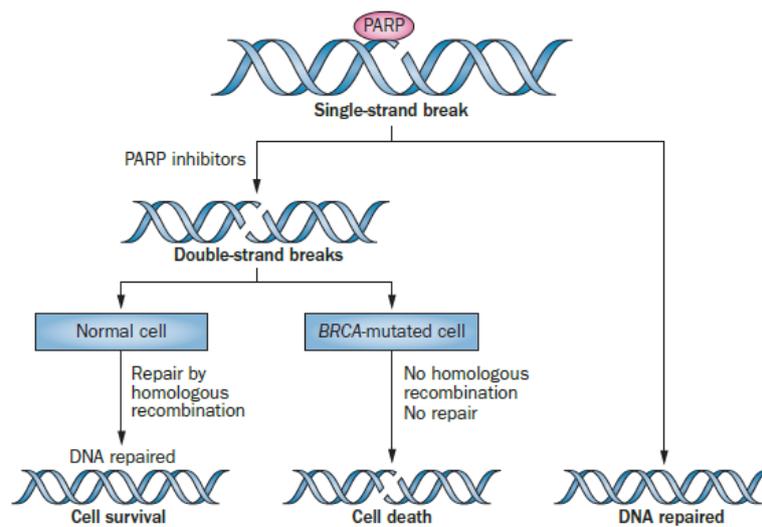


Figure 8. Le principe de létalité synthétique dans un contexte d'inhibiteurs de PARP (d'après Sonnenblick *et al.*, 2015)

I.2. Le syndrome de Lynch, prédisposition génétique au cancer colorectal

Le cancer colorectal (CCR) est le troisième cancer le plus fréquemment diagnostiqué chez l'homme et le second chez la femme, avec plus de 1,8 millions de nouveaux cas dans le monde en 2018 représentant ainsi plus de 10% des cas de cancers (Bray *et al.*, 2018 ; <https://gco.iarc.fr>). Il constitue également la deuxième cause de mortalité par cancer dans le monde en 2018 avec près de 900 000 décès. Les taux d'incidence et de mortalité du cancer colorectal varient considérablement dans le monde. En France, cela représente plus de 47 000 nouveaux cas et près de 20 000 décès pour l'année 2018 (<https://gco.iarc.fr>). Dans 60 à 65% des cas, les CCR sont d'origine sporadique, c'est-à-dire qu'ils résultent d'anomalies somatiques acquises au cours de la vie de l'individu (Figure 9 ; Keum and Giovannucci, 2019). En revanche, 35 à 40% des cas de CCR sont associés à une composante familiale, c'est-à-dire qu'un, deux ou plusieurs apparentés au premier et/ou au second degré du cas index présentent également un CCR, suggérant ainsi une contribution génétique ou l'intervention de facteurs environnementaux, ou encore, une combinaison des deux (Lichtenstein *et al.*, 2000 ; Graff *et al.*, 2017). Parmi ces formes familiales de CCR, 25% d'entre elles présentent une agrégation familiale avec une fréquence élevée de CCR mais dont la transmission n'est pas compatible avec un modèle héréditaire tandis que seulement 5% d'entre elles résultent d'une prédisposition héréditaire avérée (Figure 9 ; Jasperson *et al.*, 2010 ; Keum and Giovannucci, 2019). Parmi les prédispositions héréditaires au cancer colorectal, on peut distinguer : (i) la polypose adénomateuse familiale (PAF) transmise de façon autosomique dominante (altération constitutionnelle hétérozygote du gène *APC*) et récessive (altérations constitutionnelles bi-alléliques du gène *MUTYH*) correspondant à moins de 1% des cas de CCR et (ii) le syndrome de Lynch, anciennement dénommé HNPCC (*Hereditary Non Polyposis Colorectal Cancer*), transmis de façon autosomique dominante constituant environ 2 à 4% des cas de CCR (Baert-Desurmont *et al.*, 2018 ; Keum and Giovannucci, 2019).

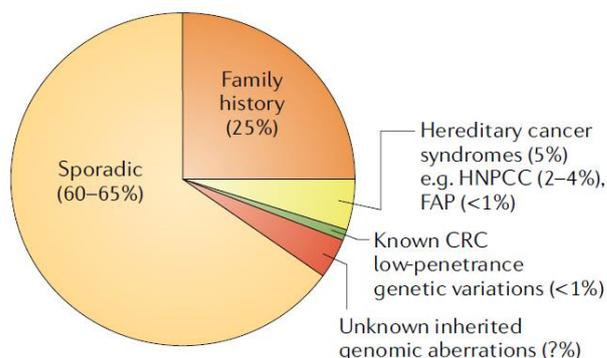


Figure 9. Proportion des différents types de cancers colorectaux (d'après Keum and Giovannucci, 2019)

Le syndrome de Lynch est l'une des plus anciennes formes décrites de prédisposition héréditaire au cancer (Lynch *et al.*, 1977 ; Lynch *et al.*, 2015). C'est en 1895 qu'Alfred Warthin, médecin américain, fait l'observation d'un très grand nombre de décès par cancer dans la famille de sa

couturière. Il rapporte dans ses travaux publiés en 1913 que la famille G, ainsi nommée du fait de son origine germanique, présente de nombreux cas de cancers colorectaux, de l'utérus et de l'estomac et ce, dans toutes les générations. Il émet alors l'hypothèse selon laquelle le cancer se transmet dans cette famille via un facteur héréditaire, avec un mode de transmission mendélien (Lynch *et al.*, 2015). Dans les années 1960, le médecin américain Henry Lynch fait l'observation de deux familles avec les mêmes traits phénotypiques que la famille G, validant ainsi l'hypothèse d'une transmission héréditaire du cancer selon un mode de transmission autosomique dominant. Ce « syndrome du cancer familial » (*cancer family syndrome*) est renommé en 1984 « syndrome de Lynch » par le médecin américain Richard Boland avant d'être appelé HNPCC par la suite (Boland and Troncale, 1984). Il a également été mis en évidence que ce syndrome prédisposait non seulement aux cancers colorectaux, mais également à un large spectre tumoral (Vasen *et al.*, 1990). En effet, ce syndrome touche majoritairement le côlon et le rectum mais il est également associé à un risque élevé de cancer de l'endomètre. En outre, il confère également des risques plus modérés de développer des tumeurs dites du spectre étroit telles que le cancer de l'intestin grêle et des voies excrétrices urinaires, mais aussi des tumeurs dites du spectre élargi telles que le cancer de l'estomac, des ovaires, du pancréas et des voies biliaires (Win *et al.*, 2012 ; Medina-Arrana *et al.*, 2012). Ainsi, la dénomination « syndrome de Lynch » est désormais préférée à celle de HNPCC. D'autres cancers héréditaires considérés comme de rares variantes de ce syndrome se caractérisent par des présentations phénotypiques particulières. Ainsi un syndrome de Muir-Torre est évoqué dans le cas d'une présentation familiale de CCR avec des tumeurs cutanées (Muir *et al.*, 1967 ; Torre *et al.*, 1968) alors qu'un syndrome de Turcot est suspecté lorsque des CCR et des tumeurs cérébrales sont identifiés au sein d'une même famille (Turcot *et al.*, 1959).

1.2.1. Les gènes *MMR*, gènes majeurs de prédisposition au syndrome de Lynch

Le syndrome de Lynch résulte d'altérations génétiques constitutionnelles hétérozygotes conduisant à la perte de fonction de l'un des gènes codant des protéines impliquées dans la voie de réparation des mésappariements de l'ADN, aussi connu sous le nom de système MMR (*MisMatch Repair*). Quatre gènes ont ainsi été identifiés comme gènes majeurs de prédisposition au syndrome de Lynch et sont appelés, dans leur ensemble, gènes *MMR*. Le premier gène identifié comme étant associé à ce syndrome a été le gène *MSH2* (*MutS Homolog 2 of E. coli* ; OMIM #609309) localisé au niveau de la région 2p21 et se composant de 16 exons codant une protéine de 934 acides aminés (Figure 10 ; Fishel *et al.*, 1993). Au cours de la même année fût découvert un second gène de prédisposition, le gène *MLH1* (*MutL homolog 1 of E. coli* ; OMIM #120436) localisé en 3p21-23 et constitué de 19 exons codant une protéine de 756 acides aminés (Figure 10 ; Lindblom *et al.*, 1993 ; Papadopoulos *et al.*,

1994). Concomitant à cette découverte, un troisième gène de prédisposition a été identifié, le gène *PMS2* (*PostMeiotic Segregation increased 2* ; OMIM #600259) localisé dans la région 7p22 et composé de 15 exons codant une protéine de 862 acides aminés (Figure 10 ; Nicolaides *et al.*, 1994). Enfin, le quatrième gène associé au syndrome de Lynch a été découvert quelques années après les trois premiers, il s'agit du gène *MSH6* (*MutS Homolog 6 of E. coli* ; OMIM #600678) localisé en 2p16 et constitué de 10 exons codant une protéine de 1 361 acides aminés (Figure 10 ; Miyaki *et al.*, 1997). De façon intéressante, il a récemment été découvert que des délétions constitutionnelles dans la partie 3' du gène *EPCAM* (*EPithelial Cell Adhesion Molecule* ; OMIM #185535), localisé au niveau de la région 2p21 en amont du gène *MSH2*, pouvaient conduire à une inactivation épigénétique de *MSH2* et constituer ainsi, une nouvelle cause de syndrome de Lynch (Ligtenberg *et al.*, 2009). L'expression des gènes *MMR* est ubiquitaire. Pourtant, le spectre tumoral associé à une défaillance d'un de ces gènes est restreint à certains tissus, comme explicité ci-dessus. A ce jour, l'origine de cette spécificité tissulaire tumorale associée à une défaillance du système *MMR* n'est pas complètement élucidée. Plusieurs hypothèses ont été avancées parmi lesquelles la vitesse de prolifération cellulaire des tissus touchés ou l'exposition à des agents environnementaux toxiques (Chao and Lipkin, 2006).

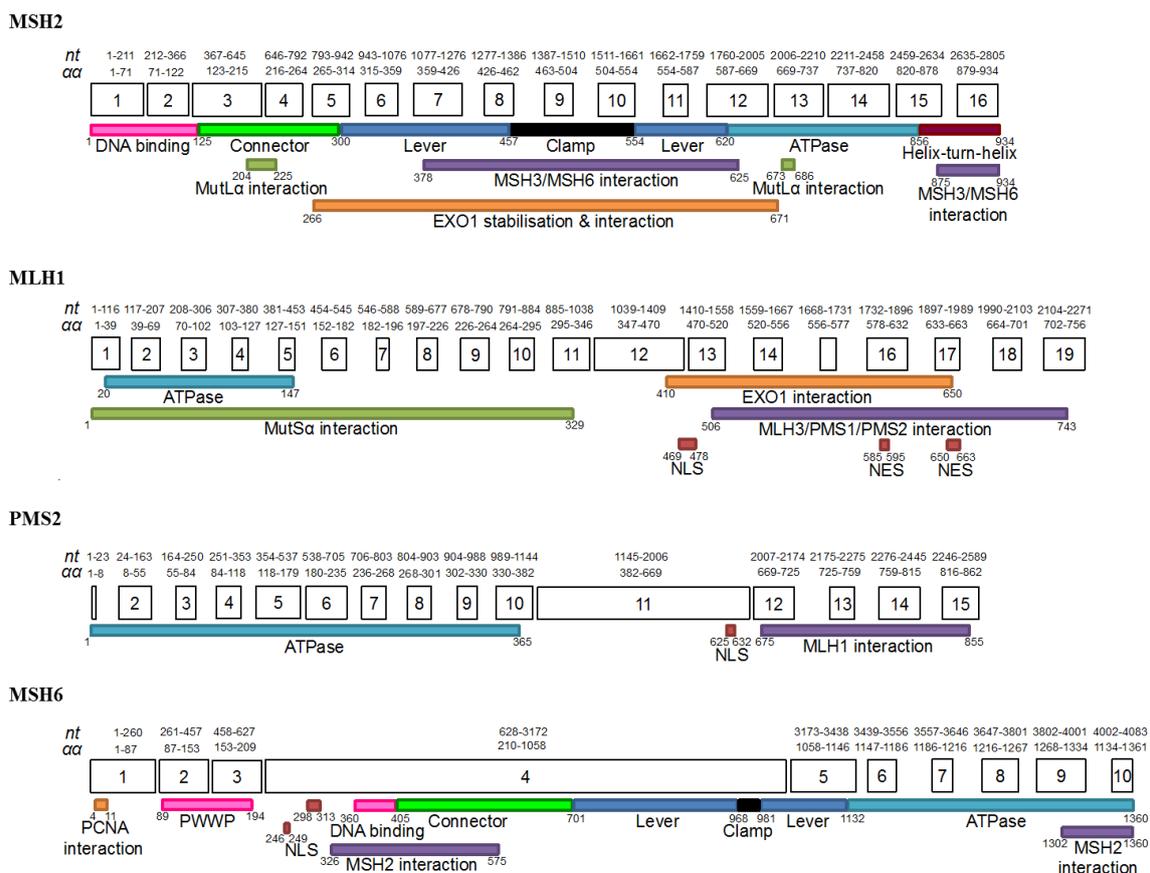


Figure 10. Représentation schématique des gènes et des domaines fonctionnels des protéines MMR (d'après les guidelines v.2.4. 2018 de l'organisation InSIGHT, https://www.insight-group.org/content/uploads/2018/08/2018-06_InSIGHT_VIC_v2.4.pdf)

Les altérations constitutionnelles hétérozygotes des gènes *MMR* sont associées à des pénétrances incomplètes variables selon le gène altéré (Moller *et al.*, 2018 ; ten Broeke *et al.*, 2018 ; Dominguez-Valentin *et al.*, 2020). Une variation à l'origine d'une perte de fonction d'un de ces gènes conduira à un risque plus ou moins important de développer une tumeur. Ainsi, les risques cumulés de développer une tumeur du spectre associé à ce syndrome à 70 ans sont estimés à 79%, 77%, 62% et 22% pour une femme et 64%, 71%, 28% et 22% pour un homme lorsque *MLH1*, *MSH2*, *MSH6* ou *PMS2* sont mutés, respectivement (Tableau 3 ; Dominguez-Valentin *et al.*, 2020). Cette variabilité d'incidence entre les hommes et les femmes peut s'expliquer par le fait que le spectre du syndrome de Lynch est caractérisé en grande partie par des cancers touchant des organes spécifiquement féminins tels que l'endomètre ou encore, les ovaires. L'incidence, en plus de varier selon le gène porteur de l'anomalie pathogène, peut donc également varier selon la nature du cancer ou selon le sexe de l'individu. En effet, le risque cumulé d'atteinte d'un cancer colorectal à l'âge de 70 ans est de 44%, 42%, 20% et 3% chez les femmes porteuses d'une variation délétère de *MLH1*, *MSH2*, *MSH6* ou *PMS2*, respectivement, tandis que les individus masculins présenteront un risque cumulé de 53%, 46%, 12% ou 3%, respectivement (Tableau 3 ; Dominguez-Valentin *et al.*, 2020). Après le cancer colorectal, le cancer de l'endomètre est le second cancer du spectre à présenter l'incidence la plus élevée avec un risque cumulé de développer une tumeur à 70 ans de 35%, 47%, 41% et 13% pour les femmes porteuses d'une variation causale de *MLH1*, *MSH2*, *MSH6* et *PMS2* (Tableau 3 ; Dominguez-Valentin *et al.*, 2020).

Organ	<i>MLH1</i>		<i>MSH2</i>		<i>MSH6</i>		<i>PMS2</i>
	Females	Males	Females	Males	Females	Males	Both
Any cancer	77.8 [71.1–84.5]	63.5 [55.9–72.1]	76.9 [69.3–84.3]	70.6 [61.4–81.3]	61.8 [47.3–78.3]	27.9 [16.2–48.7]	21.7 [11.1–48.2]
Colorectal	44.1 [37.4–51.8]	52.8 [45.2–61.6]	41.9 [34.9–49.7]	46.3 [36.9–58.8]	20.3 [11.8–40.5]	11.7 [4.7–35.2]	3.4 [0.6–34.5]
Endometrium	35.2 [28.8–43.4]		46.5 [38.3–56.3]		41.1 [28.6–57.9]		12.8 [5.2–49.5]
Ovaries	11.0 [7.4–17.1]		17.4 [11.8–27.3]		10.8 [3.7–32.9]		3.0 [0.5–43.3]
Stomach, small bowel, bile duct, gallbladder and pancreas	8.4 [5.5–13.0]	15.7 [11.2–22.0]	9.8 [6.7–14.5]	15.9 [11.3–22.4]	1.7 [0.3–23.8]	4.0 [1.1–26.2]	3.6 [1.0–33.5]
Ureter and kidney	2.9 [1.4–6.3]	3.7 [1.9–7.9]	13.3 [9.4–18.8]	16.2 [11.6–22.6]	5.5 [2.2–26.9]	1.7 [0.3–24.3]	0 [0–30.9]
Urinary bladder	2.7 [1.2–6.2]	4.6 [2.5–8.7]	6.8 [4.2–11.3]	8.7 [5.5–13.8]	1.2 [0.2–23.4]	4.3 [1.2–26.5]	0 [0–30.9]
Prostate	7.0 [4.2–11.9]		15.9 [11.2–22.5]		4.8 [1.3–27.0]	4.6 [0.8–67.5]	
Breast	10.5 [7.4–15.1]		12.6 [8.9–17.8]		11.1 [5.8–31.5]		8.1 [2.8–46.5]
Brain	1.6 [0.6–4.6]	0.7 [0.1–4.2]	1.8 [0.7–5.4]	3.7 [1.8–8.4]	1.2 [0.2–23.4]	1.8 [0.3–24.4]	0 [0–30.9]

Tableau 3. Risques cumulés à 70 ans de développer une tumeur du spectre du syndrome de Lynch en fonction du gène altéré (adapté de Dominguez-Valentin *et al.*, 2020). Les chiffres sont exprimés en pourcentage.

Il est d'ailleurs intéressant de noter que certaines tumeurs sont préférentiellement associées à une altération d'un gène donné du système MMR (Tableau 4 ; Lynch *et al.*, 2015). Par exemple, les CCR se développent essentiellement dans le cas de variations délétères du gène *MLH1* tandis qu'une présentation de cancers extracoliques sera plus largement associée à une altération causale du gène *MSH2* (Peltomäki *et al.*, 2001). De même, les cancers de l'endomètre résultent majoritairement de variations pathogènes dans le gène *MSH6* (Hendriks *et al.*, 2004). Néanmoins, la grande majorité des cas de syndrome de Lynch sont associés à des variations constitutionnelles hétérozygotes pathogènes identifiées dans les gènes *MSH2* et *MLH1*. En effet, selon le rapport de l'institut national du cancer (INCa) de 2017, 36% des variations détectées dans un cas de syndrome de Lynch sont localisées dans le gène *MSH2*, 28% dans le gène *MLH1*, 25% dans le gène *MSH6* et enfin 11% dans le gène *PMS2* (Buecher *et al.*, 2019 ; Consultations, laboratoires et suivi, INCa, 2017). De façon intéressante, une étude américaine effectuée en 2015 sur une cohorte de 1 260 individus tend à démontrer un ratio de variations identifiées par gène dans des cas de syndrome de Lynch très similaire au rapport de l'INCa tel que 35% sont des variations localisées dans le gène *MSH2*, 27% dans *MLH1*, 23% dans *MSH6*, 12% dans *PMS2* et 3% dans *EPCAM* (Yurgelun *et al.*, 2015). Plus rarement, certaines tumeurs associées au syndrome de Lynch peuvent être caractérisées par une hyperméthylation du promoteur du gène *MLH1* conduisant ainsi à l'inhibition de la transcription de l'allèle muté (Peltomäki *et al.*, 2014). Toutefois, il est possible que ces données soient biaisées du fait de la différence de pénétrances associées à ces différents gènes. En effet, les formes de syndrome de Lynch résultant d'altérations des gènes *MLH1* et *MSH2* sont les plus sévères et sont donc plus facilement identifiables.

Causative mechanism	Associated phenotypic heterogeneity
Heterozygous <i>MLH1</i> mutation	LS: CRC predominance; extracolonic cancers less frequent than with <i>MSH2</i> mutations
Heterozygous <i>MSH2</i> mutation	LS: greater frequency of extracolonic cancers
Heterozygous <i>MSH6</i> mutation	LS: predominance of endometrial cancer; tumours sometimes exhibit low-level MSI
Heterozygous <i>PMS2</i> mutation	LS: may contain excess colonic polyps; lower frequency of cancer
Heterozygous <i>EPCAM</i> deletion	LS: silences <i>MSH2</i> expression; often lower risk of extracolonic cancers, although if the deletion is close to the <i>MSH2</i> gene, risk for endometrial cancer increases
Monoallelic <i>MLH1</i> epimutation	LS: phenotypic expression seems to be similar to that of <i>MLH1</i> mutation carriers; a proportion of cases with <i>MLH1</i> epimutation are inherited but, more commonly, this epigenetic defect arises <i>de novo</i>

Tableau 4. L'hétérogénéité phénotypique du syndrome de Lynch est en partie fonction de la nature du gène MMR altéré (d'après Lynch *et al.*, 2015). LS, Lynch Syndrome ; CRC, ColoRectal Cancer.

Les protéines codées par les gènes *MMR* sont impliquées dans l'une des voies de réparation de l'ADN et assurent de ce fait, le maintien de l'intégrité du génome. A ce titre, les gènes *MMR*, comme les gènes *BRCA*, sont considérés comme des gènes suppresseurs de tumeurs et répondent donc au modèle « *two-hit* » de Knudson dans lequel une inactivation des deux allèles de l'un des gènes *MMR*, la première étant héritée et la seconde acquise au cours de la vie, est nécessaire pour conduire au développement tumoral (Peltomäki *et al.*, 2014). A la suite de l'inactivation des deux allèles, le système MMR normalement en charge de la réparation des erreurs de mésappariements de l'ADN au cours de la réplication, devient inactif provoquant ainsi une accumulation des erreurs dans la séquence de l'ADN avec pour conséquence une augmentation drastique du taux de mutations dans les cellules déficientes par rapport aux cellules normales. Cette instabilité génomique est à l'origine du processus de tumorigenèse. L'inactivation du système MMR entraîne l'accumulation de mutations somatiques dans plusieurs gènes suppresseurs de tumeurs, tels que *TGFBR2*, *BAX*, *APC* et *PTEN*, induisant de ce fait une rapide progression de la tumeur (The Cancer Genome Atlas Network, 2012).

1.2.2. Les fonctions associées aux protéines MMR

Les gènes *MMR* codent des protéines aux fonctions multiples principalement connues pour leur rôle capital dans le maintien de l'intégrité de l'information génétique au cours des divisions cellulaires. Les fonctions essentielles du système MMR sont, d'une part, la réparation post-réplivative des mésappariements de l'ADN et, d'autre part, la signalisation apoptotique de certains dommages à l'ADN, mais elles interviennent également dans diverses fonctions cellulaires (Hsieh and Yamane, 2008 ; Edelbrock *et al.*, 2013 ; Kunkel and Erie, 2015).

1.2.2.a. La réparation post-réplivative des mésappariements de l'ADN

Comme explicité dans la partie 1.1.2.a. traitant de la fonction des protéines BRCA dans le système de réparation des cassures double brin par recombinaison homologue, la réplication de l'ADN est une étape charnière dans la transmission de l'information génétique aux cellules filles. Il est indispensable que l'intégrité de cette information soit maintenue au plus proche de l'information contenue dans l'ADN de la cellule mère. Le système MMR est un mécanisme bidirectionnel de réparation des mésappariements de l'ADN qui permet de corriger les erreurs survenues lors de la réplication de l'ADN ayant échappé au contrôle des polymérase (*proofreading*) (Hsieh and Yamane, 2008 ; Edelbrock *et al.*, 2013 ; Kunkel and Erie, 2015 ; Liu *et al.*, 2017). Cette voie de réparation permet de détecter et d'exciser précisément le mésappariement avant de resynthétiser correctement le brin néoformé. Grâce à ce système, le taux d'erreurs de la réplication diminue d'un facteur 1 000,

aboutissant à moins d'une erreur par génome diploïde lors de la division d'une cellule. C'est en 1975 que l'équipe du biologiste et généticien américain Matthew Meselson a découvert cette voie de réparation dans la bactérie *Escherichia coli* (*E. coli*) (Wildenberg and Meselson, 1975 ; Wagner and Meselson). C'est ensuite l'équipe du biochimiste américain Paul Modrich qui a pleinement caractérisé le système MMR de *E. coli*, alors appelé système MutHLS (Mut, *Mutator*), en le reconstituant *in vitro* et en isolant les 11 protéines responsables de cette activité (Su and Modrich, 1986 ; Su *et al.*, 1988). Par la suite, l'équipe de Paul Modrich a démontré qu'un système analogue existait dans les cellules eucaryotes (Fang and Modrich, 1993 ; Modrich, 2006). En effet, cette voie de réparation est très conservée de la bactérie aux mammifères ce qui se traduit par de très fortes homologues entre les systèmes MMR de la bactérie, de la levure et de l'homme (Eisein, 1998 ; Aravind *et al.*, 1999 ; Fukui, 2010). Cette découverte fondamentale a été récompensée par un prix Nobel de Chimie en 2015.

Le système MMR met en jeu un grand nombre de protéines, souvent associées sous la forme d'hétérodimères, remplissant des rôles spécifiques et coordonnés dans ce processus complexe et séquentiel. Chez l'homme, deux principaux complexes peuvent être distingués : (i) le complexe MutS qui est à l'origine de la reconnaissance du mésappariement et (ii) le complexe MutL qui sert d'interface entre le premier complexe et les acteurs de la réparation de l'erreur de réplication. Selon le type de mésappariements de l'ADN, le complexe MutS recruté ne sera pas constitué des mêmes protéines. En effet, l'hétérodimère MutS α , composé des protéines MSH2 et MSH6, reconnaît principalement les erreurs de type substitutions ponctuelles et petites insertions/délétions tandis que l'hétérodimère MutS β , formé des protéines MSH2 et MSH3, reconnaît exclusivement de larges insertions/délétions. Il est à noter que le complexe MSH2-MSH6 constitue l'hétérodimère MutS le plus abondant et à ce titre, reconnaît la majorité des mésappariements de l'ADN (Reyes *et al.*, 2015). Le complexe MutL, quant à lui, peut se former à partir des protéines MLH1 et PMS2, constituant ainsi l'hétérodimère MutL α , le plus connu chez l'homme, mais il peut également se composer des protéines MLH1-PMS1 et MLH1-MSH3 formant, respectivement, les hétérodimères MutL β et MutL γ dont les rôles dans ce mécanisme sont mineurs. Les protéines MLH1, MSH2, MSH6 et PMS2 exercent donc des fonctions essentielles dans le système MMR. La majorité des cas de syndrome de Lynch est associée à une altération pathogène des gènes *MLH1* et *MSH2* du fait de l'implication systématique des deux protéines résultantes dans la formation des hétéroduplexes MutS et MutL, respectivement (Plazzer *et al.*, 2013).

La réparation des mésappariements de l'ADN par le système MMR est initiée avec la reconnaissance de l'erreur de réplication par l'hétérodimère MutS α (Figure 11). L'interaction avec le mésappariement induit un changement de conformation du complexe MutS α qui, après liaison à l'ATP, recrute l'hétérodimère MutL α formant ainsi un complexe tétramérique. L'activité endonucléase

intrinsèque de MutL α est ensuite activée de manière stochastique par la protéine PCNA (*Proliferating Cell Nuclear Antigen*), chargée elle-même sur l'ADN par le facteur RFC (*Replication Factor C*), afin d'inciser le brin néosynthétisé, porteur du mésappariement (Figure 11 ; Reyes *et al.*, 2015 ; Kunkel and Erie, 2015 ; Liu *et al.*, 2017). Le mécanisme de discrimination du brin porteur du mésappariement n'est pas encore bien défini mais il a été suggéré que ce serait la liaison asymétrique de la protéine PCNA avec le brin néoformé qui permettrait à l'hétérodimère MutL α d'inciser spécifiquement le brin néosynthétisé (Kadyrov *et al.*, 2006). Cependant, il a également été proposé que MutL α soit en capacité d'identifier précisément l'un des brins néoformés du fait de sa nature discontinue. En effet, la réplication bidirectionnelle de l'ADN est intrinsèquement asymétrique conduisant ainsi à la production à la fois de segments longs sur un brin (*leading strand*) et de courts segments discontinus transitoires nommés fragments d'Okazaki sur l'autre brin (*lagging strand*) (Pavlov *et al.*, 2003). Par la suite, les protéines MSH2 et MLH1 du complexe tétramérique vont recruter l'enzyme exonucléase 1 (EXO1) qui va se servir de l'incision réalisée par l'hétérodimère MutL α comme point d'ancrage afin d'éliminer une partie du brin néosynthétisé (Figure 11). Il apparaît qu'une déficience en cette enzyme ne conduit pas nécessairement à une inactivation du système MMR, suggérant ainsi l'existence de deux voies au sein de ce système : l'une, EXO1-dépendante et l'autre, EXO1-indépendante (Jagmohan-Changur *et al.*, 2003). Plusieurs hypothèses supportent l'existence de cette seconde voie dont, notamment, celle considérant MRE11, une exonucléase généralement associée au mécanisme de réparation des cassures double brin de l'ADN. En effet, il a été montré qu'une déficience en MRE11 ou une anomalie dans son domaine d'interaction avec la protéine MLH1, sont associées à une déficience du système MMR (Vo *et al.*, 2005). Pendant la dégradation du brin néosynthétisé, des protéines RPA sont recrutées au niveau du brin d'ADN qui sert de matrice afin de le stabiliser. Enfin, une fois la dégradation du brin porteur du mésappariement effectuée, les ADN polymérases δ et ϵ pourront re-synthétiser fidèlement le brin d'ADN à partir de la séquence d'ADN complémentaire du brin matrice et l'ADN ainsi nouvellement synthétisé sera scellé au reste du brin grâce à une ligase d'ADN (Figure 11).

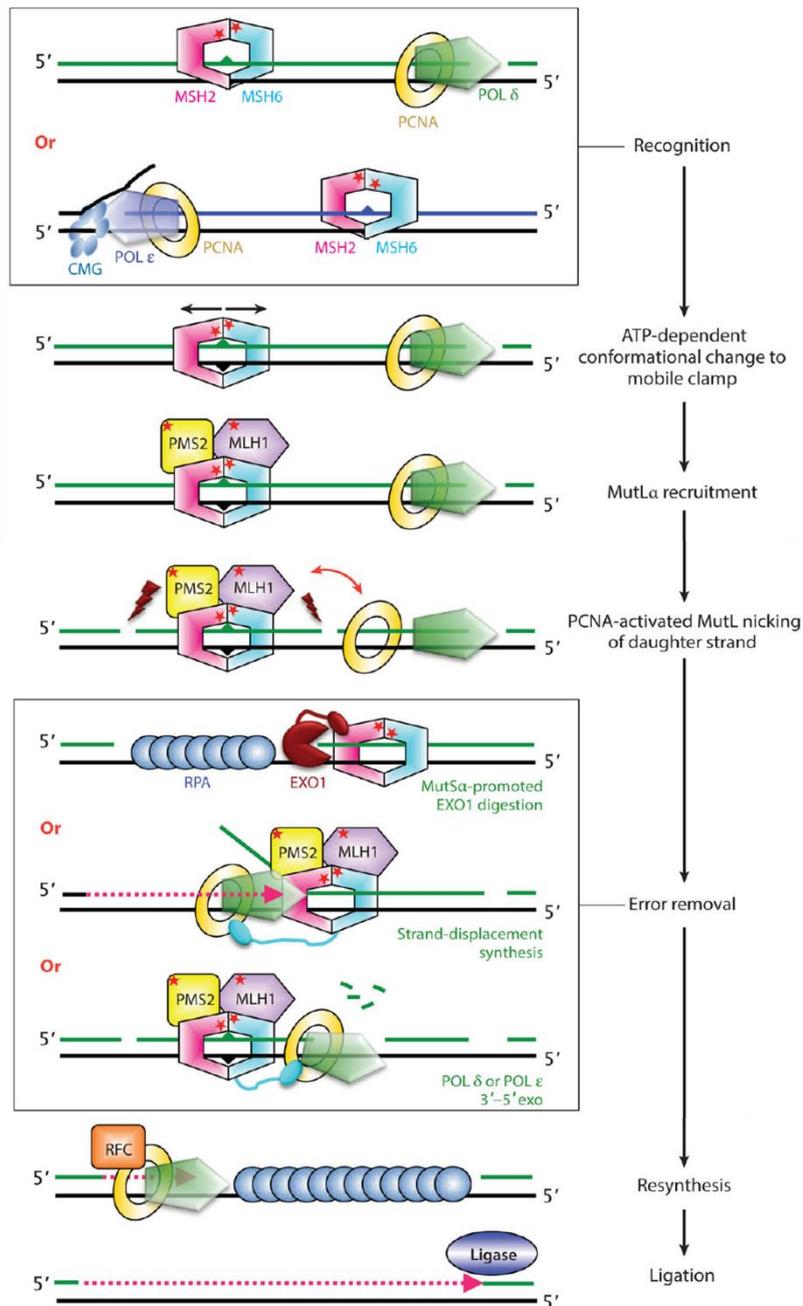


Figure 11. Le mécanisme de réparation post-répliquative des mésappariements de l'ADN par le système MMR chez l'homme (d'après Kunkel and Erie, 2015)

1.2.2.b. La signalisation apoptotique de certains dommages de l'ADN

En plus de sa fonction de réparation des mésappariements de l'ADN, le système MMR intervient également dans la réponse cellulaire à certains agents génotoxiques. En effet, les hétéroduplexes MutS sont capables de reconnaître des altérations de l'ADN induites par des agents chimiques de type antinéoplasiques ou alkylants, conditionnant ainsi l'arrêt du cycle cellulaire et l'apoptose de la cellule (Jiricny, 2006 ; O'Brien and Brown, 2006 ; Hsieh and Yamane, 2008 ; Gupta and

Heinen, 2019). De façon intéressante, il a été montré que les cellules déficientes en protéines MMR exposées à des agents alkylants étaient 100 fois plus résistantes à la mort cellulaire illustrant ainsi le rôle important de ces protéines dans la signalisation apoptotique en réponse à des agents alkylants (Karran, 2001).

Les agents alkylants MNNG (N-méthyl-N'-nitro-N-nitrosoguanidine), MNU (N-méthyl-N-nitrosourée) et leurs analogues utilisés comme agents chimiothérapeutiques anticancéreux, le témozolomide et la dacarbazine, sont à l'origine d'une cytotoxicité cellulaire du fait de la méthylation de l'atome d'oxygène en position 6 des guanines de l'ADN (O⁶-meG, O⁶-méthylGuanine) (Swan *et al.*, 1996 ; Karran, 2001). Classiquement, l'enzyme MGMT (O⁶-MéthylGuanine-Méthyltransférase) constitue la première défense contre les lésions O⁶-meG par transfert du groupement méthyle de la guanine altérée vers un de ses propres résidus cystéine, ce qui a pour conséquence de l'inactiver (Yarosh *et al.*, 1984). Néanmoins, cette enzyme n'est pas infaillible et peut se retrouver saturée ou inactivée induisant ainsi une fuite du premier système de réparation de ce type d'altérations. Pendant la réplication, la base modifiée O⁶-meG n'est pas correctement reconnue par la polymérase qui va incorporer une thymine au lieu d'une cytosine, base complémentaire de la guanine, générant un mésappariement. L'hétéroduplexe MutS reconnaît alors le mésappariement apparu sur le brin néosynthétisé et induit l'apoptose (Duckett *et al.*, 1996 ; Karran, 2001 ; Mojas *et al.*, 2007).

Aujourd'hui, le mécanisme à l'origine de la mort cellulaire à la suite de la reconnaissance d'un mésappariement O⁶-meG par le système MMR n'est pas complètement élucidé et deux modèles non exclusifs sont généralement discutés (Figure 12 ; Edelbrock *et al.*, 2013 ; Gupta and Heinen, 2019). Dans le premier modèle appelé « modèle de signalisation directe », il est proposé que le système MMR soit directement à l'origine de l'activation de la voie de signalisation des dommages à l'ADN par interaction des protéines MSH2 et ATR (*Ataxia-Telangiectasia mutated and Rad3-Related*) et/ou de la protéine MLH1 avec la protéine ATM (Brown *et al.*, 2003 ; Wang and Qin, 2003 ; Yoshioka *et al.*, 2006 ; Jiricny, 2006). L'arrêt du cycle cellulaire serait médié par les protéines CHK1 et CHK2 (*CHeckpoint Kinase 1/2*) à la suite de leur interaction avec les protéines ATR et ATM, ou encore directement par liaison avec la protéine MSH2 (Hawn *et al.*, 1995 ; Adamson *et al.*, 2005). Dans ce modèle, il semblerait que l'apoptose soit provoquée par phosphorylation de la protéine p53 (Duckett *et al.*, 1999 ; Hickman and Samson, 1999). Le second modèle, nommé « modèle de réparation futile » ou « modèle des cycles futiles », suggère que ce sont les multiples tentatives de réparation du mésappariement par le système MMR qui génèrent des portions d'ADN simple brin puis double brin, au fur et à mesure des cycles de réplication, conduisant à terme à l'arrêt du cycle cellulaire et à l'apoptose (Figure 12 ; York and

Modrich, 2006 ; Mojas *et al.*, 2007). Dans ce modèle, le système MMR ne peut réparer la véritable lésion car celle-ci est localisée sur le brin matrice et pas sur le brin néosynthétisé.

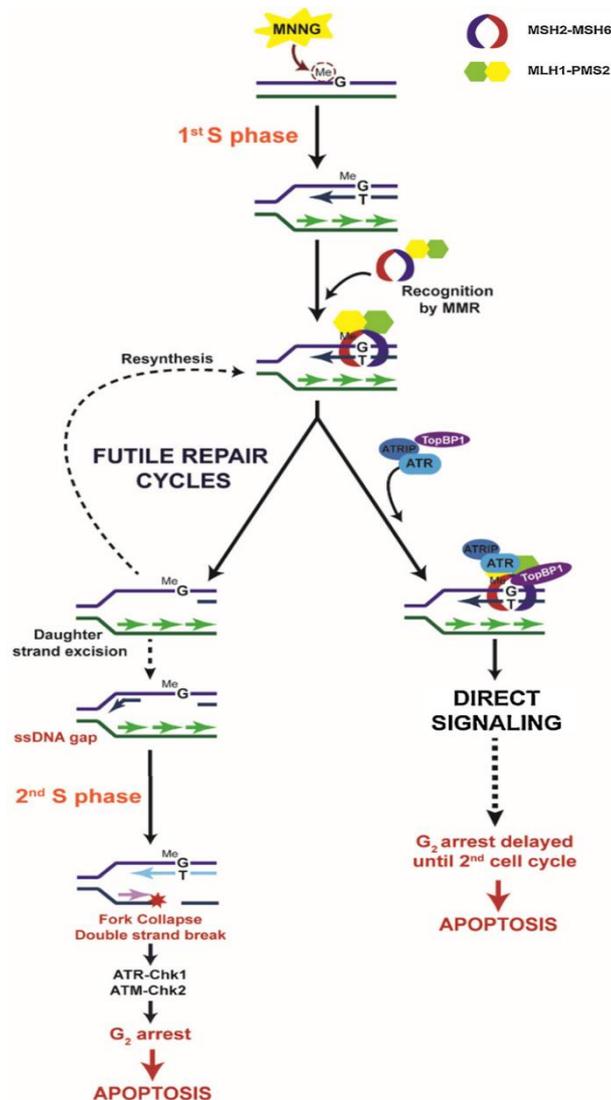


Figure 12. Les deux modèles d'induction de l'apoptose à la suite de la reconnaissance d'un mésappariement O⁶-meG par le système MMR (adapté de Gupta and Heinen, 2019)

Il est à noter que le système MMR est également impliqué dans la réponse cellulaire aux lésions oxydatives de type 8-oxoguanine (Colussi *et al.*, 2002 ; Mazurek *et al.*, 2002) ou encore, aux dommages induits pas des rayonnements ultraviolets (Young *et al.*, 2004 ; van Oosten *et al.*, 2005). Ainsi, les protéines majeures du système MMR, en plus de leur rôle au sein de la voie de réparation des mésappariements de l'ADN, modulent plusieurs mécanismes de contrôle de l'intégrité du génome. La déficience d'une de ces protéines conduit non seulement à l'accumulation d'anomalies mais aussi à la croissance sélective de cellules mutées, conduisant ainsi à la tumorigenèse.

1.2.2.c. Les autres fonctions portées par les protéines MMR

Outre la correction des mésappariements de l'ADN lors de la réplication et la signalisation apoptotique de certains dommages à l'ADN, les protéines MMR interviennent également au sein d'autres voies de réparation de l'ADN (Jiricny, 2006 ; Gupta and Heinen, 2019). En effet, toute altération de la conformation hélicoïdale de l'ADN constitue une cible potentielle pour les composants du système MMR. Par exemple, au cours de la réparation des cassures double brin de l'ADN par recombinaison homologue, il est essentiel que le chromosome porteur de l'anomalie recombine spécifiquement avec la séquence homologue associée. Dans le cas d'une recombinaison hétérologue, la réunion de deux portions d'ADN ne présentant pas ou peu d'homologie de séquence conduira à l'introduction de mésappariements (George and Alani, 2012 ; Chakraborty and Alani, 2016). Les hétéroduplexes MutS α ou β , dépendants du type de mésappariements, sont en capacité d'identifier ces anomalies, de s'y fixer et de recruter un complexe protéique spécifique permettant le déroulement de la molécule d'ADN et ce, afin d'initier une nouvelle recherche d'homologie de séquence. Ce processus, appelé anti-recombinaison, joue un rôle indispensable dans le maintien de l'intégrité du génome.

De plus, certaines protéines MMR sont amenées à jouer un rôle dans diverses fonctions cellulaires telles que la recombinaison mitotique (Datta *et al.*, 1996 ; Rizki and Lundblad, 2001) et méiotique (Hunter and Borts, 1997 ; Wang *et al.*, 1999b) ainsi que la commutation de classes des immunoglobulines (Ehrenstein and Neuberger, 1999 ; Schrader *et al.*, 1999).

1.2.2.d. Les domaines fonctionnels des protéines MMR

Comme illustré dans la figure 10, les protéines majeures du système MMR présentent des domaines fonctionnels spécifiques, parfois chevauchants, importants pour leurs activités de réparation de l'ADN. Les protéines MSH2 et MSH6 portent des domaines d'interaction permettant la formation de l'hétérodimère MutS α (Edelbrock *et al.*, 2013). De plus, ces protéines possèdent chacune un domaine de liaison à l'ADN (DBD, *DNA Binding Domain*) conditionnant la fixation de l'hétérodimère MutS α au brin d'ADN porteur du mésappariement. Le domaine « connecteur » (*connector*), comme sa dénomination l'indique, permet la connexion entre le brin d'ADN et une troisième région, le domaine « levier » (*lever*) séparé en deux zones distinctes par le domaine « pince » (*clamp*). La combinaison de ces trois domaines assure le maintien de l'hétérodimère sur l'ADN endommagé favorisant ainsi l'activation du domaine ATPase, localisé en C-terminal des deux protéines (Edelbrock *et al.*, 2013). La protéine MSH6 contient également un motif PWWP (Pro-Trp-Trp-Pro), un signal de localisation

nucléaire (NLS, *Nuclear Localisation Signal*) et un domaine d'interaction N-terminal avec la protéine PCNA (PIP, *PCNA-Interacting Protein*). La protéine MSH2 est aussi constituée d'un domaine de fixation à MSH3 permettant la formation de l'hétérodimère MutS β , d'un motif hélice-tour-hélice (HTH, *Helix-Turn-Helix*), d'un domaine d'interaction et de stabilisation à EXO1 et d'un domaine de liaison à l'hétérodimère MutL α . Ce dernier est formé des protéines MLH1 et PMS2, toutes deux constituées, en C-terminal, d'un domaine de liaison à son partenaire (Kadyrova and Kadyrov, 2016). Elles possèdent également un domaine NLS ainsi qu'un domaine ATPase situé en N-terminal, nécessaire à l'activité endonucléase du complexe MutL α . Cette activité est en partie assurée par le motif DQHA(X)₂E(X)₄E, porté par la région C-terminale de la protéine PMS2, mais également PMS1 et MLH3 (Kadyrov *et al.*, 2006). La protéine MLH1 contient également un domaine de liaison à l'hétérodimère MutS α , un domaine d'interaction avec EXO1, plusieurs signaux d'exportation nucléaire (NES, *Nuclear Export Signal*) ainsi qu'un domaine de fixation aux protéines PMS1 et MLH3 permettant la formation des hétérodimères MutL β et MutL γ , respectivement.

1.2.3. Le diagnostic du syndrome de Lynch

Le diagnostic définitif d'un syndrome de Lynch repose sur l'identification, dans le génome du patient, d'une altération génétique constitutionnelle hétérozygote à l'origine de la perte de fonction de l'un des gènes *MMR* (Lynch *et al.*, 2015 ; Yurgelun and Hampel, 2018). Ce diagnostic moléculaire n'est cependant réalisé qu'après l'identification en amont d'un certain nombre d'éléments évocateurs de la maladie tels que le contexte clinique individuel, les antécédents familiaux et les caractéristiques tumorales.

1.2.3.a. Identification des cas évocateurs

Une prédisposition héréditaire au cancer est évoquée devant la précocité de survenue de la tumeur, une multiplicité des cancers chez un même individu ainsi qu'une agrégation de CCR et/ou de cancers du spectre du syndrome de Lynch au sein d'une même branche familiale (Figure 13). Dans ce contexte, seront entreprises des analyses à la recherche d'une déficience du système MMR au niveau tumoral. Celles-ci comprennent, d'une part, l'évaluation de l'instabilité microsatellitaire (MSI, *MicroSatellite Instability*) au niveau de l'ADN tumoral et, d'autre part, l'analyse de l'expression somatique des protéines MMR.

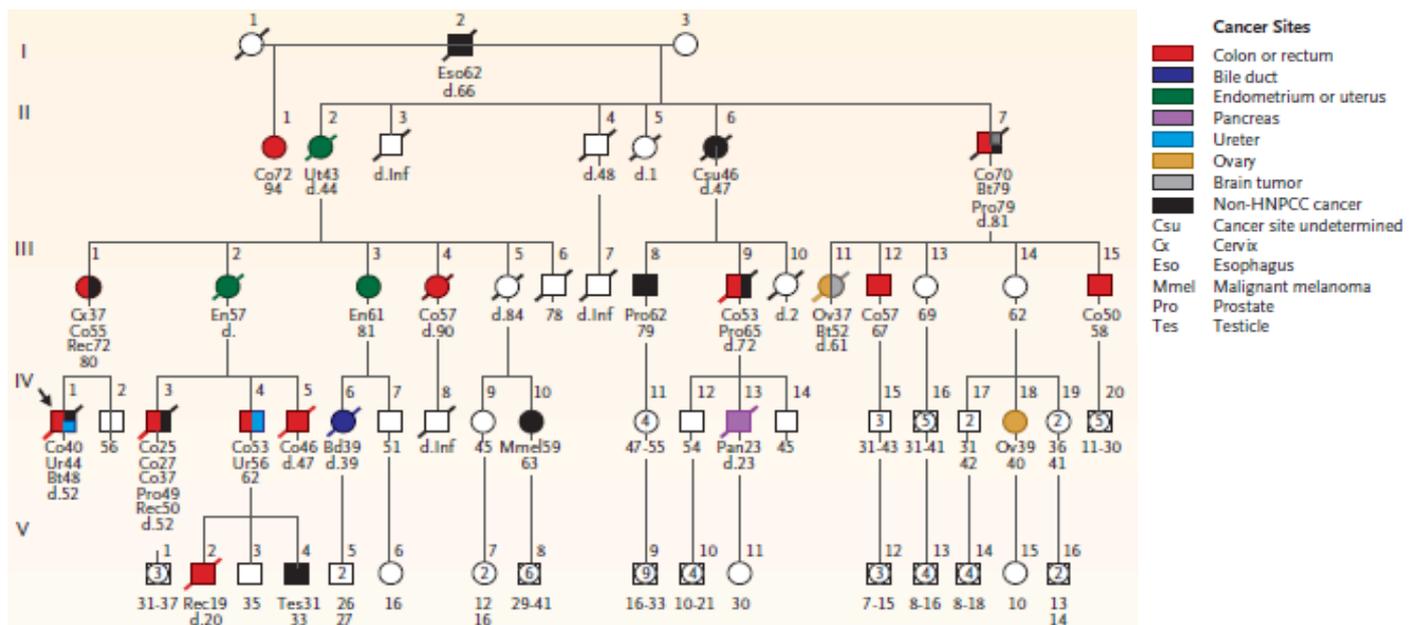


Figure 13. Arbre généalogique d'une famille évocatrice d'un syndrome de Lynch (d'après Lynch and de la Chapelle, 2013)

Évaluation du statut d'instabilité microsatellitaire

Décrit pour la première fois en 1993, les tumeurs associées à un syndrome de Lynch sont spécifiquement caractérisées par un phénotype d'instabilité microsatellitaire, signature moléculaire d'une conséquence directe de la déficience du système MMR (Thibodeau *et al.*, 1993 ; Peltomaki *et al.*, 1993). En effet, lorsqu'une des protéines indispensables au bon fonctionnement du système MMR est altérée, les erreurs accumulées lors de la réplication de l'ADN conduisent à une augmentation drastique de l'instabilité génomique tout particulièrement au niveau des microsatellites, séquences d'ADN mono, di, tri ou tétra-nucléotidiques répétées en tandem. L'ADN issu des cellules tumorales sera donc particulièrement marqué d'insertions et de délétions dans les séquences répétées des microsatellites en cas de déficience du système MMR, en comparaison à l'ADN issu des cellules du tissu normal (Sameer *et al.*, 2014). Un criblage par PCR basé sur l'amplification spécifique d'au moins cinq marqueurs mono-nucléotidiques parmi les sept recommandés (BAT25, BAT26, BAT40, NR21, NR22, NR24, NR27) est nécessaire afin de déterminer le degré d'instabilité des microsatellites dans la tumeur (Figure 14 ; Manceau *et al.*, 2011 ; Pagin *et al.*, 2013 ; Richman *et al.*, 2015 ; INCa, 2016). Trois phénotypes peuvent être observés : (i) le phénotype MSS (*MicroSatellite Stable*), lorsqu'aucune altération n'est détectée dans les marqueurs microsatellitaires, (ii) le phénotype MSI-L (*MicroSatellite Instability Low*), lorsqu'un marqueur sur les cinq est associé à une instabilité microsatellitaire et (iii) le phénotype MSI-H (*MicroSatellite Instability High*), lorsqu'au moins deux marqueurs sur les cinq présentent une instabilité microsatellitaire. Les résultats de ce type d'analyse sont toutefois à nuancer

car, dans environ 15% des cas de CCR sporadiques, un phénotype MSI peut également être détecté (Richman *et al.*, 2015). Dans ce contexte, l'instabilité microsatellitaire résulte essentiellement au niveau somatique d'une hyperméthylation du promoteur du gène *MLH1* à l'origine de la perte de fonction de la protéine. De plus, les CCRs sporadiques présentent dans une majorité des cas une variation dans l'oncogène *BRAF* (le plus souvent p.V600E), ce qui n'est pas le cas des CCRs identifiés dans un contexte de syndrome de Lynch (Domingo *et al.*, 2005). Ainsi, une instabilité microsatellitaire associée à une hyperméthylation du promoteur du gène *MLH1* et à la présence de la variation p.V600E du gène *BRAF*, sont des arguments en faveur d'une présentation de CCR sporadique, excluant un diagnostic de syndrome de Lynch.

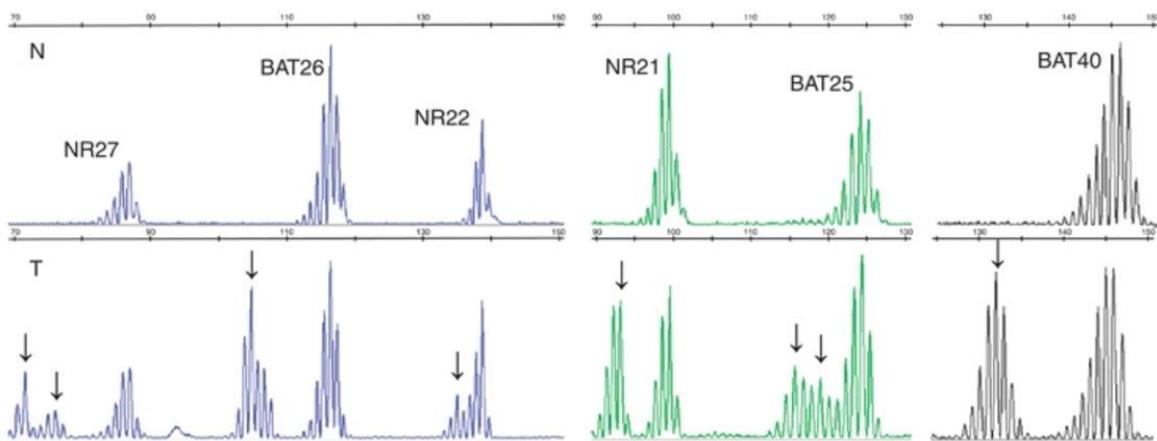


Figure 14. Profil d'instabilité microsatellitaire tumorale obtenu par PCR multiplex fluorescente suivi d'une électrophorèse capillaire sur séquenceur (d'après Pagin *et al.*, 2013). L'analyse présentée en haut correspond à l'expression microsatellitaire d'un tissu normal (N) pour les 6 marqueurs étudiés ici. L'analyse présentée en bas correspond à celle du tissu tumoral (T) colorectal d'un patient porteur d'une variation constitutionnelle du gène *MLH1* dont l'expression protéique a été perdue dans la tumeur. Cette figure montre une instabilité des microsatellites avec un nombre de répétitions altéré dans le tissu tumoral par rapport au tissu normal.

Analyse de l'expression des protéines MMR

Afin d'identifier toujours plus précisément les patients susceptibles de présenter une altération constitutionnelle de l'un des gènes *MMR* et donc un syndrome de Lynch, une analyse de l'expression protéique des composants principaux du système MMR dans les tissus tumoraux est généralement recommandée. Dans la majorité des cas, les tumeurs déficientes au niveau du système MMR présentent une extinction totale d'expression d'au moins une des protéines MMR étant donné qu'une altération dans un des gènes *MMR* conduit souvent à la perte d'expression de la protéine associée. L'immunohistochimie (IHC), méthode rapide et simple, permet d'interroger directement l'expression des quatre protéines dans le tissu tumoral et ainsi, d'orienter l'étude moléculaire

constitutionnelle (Figure 15 ; Thibodeau *et al.*, 1996 ; Yurgelun and Hamel, 2018). Une perte d'expression des protéines MSH6 ou PMS2 suggère la présence d'une altération constitutionnelle dans l'un des gènes associés tandis qu'un défaut d'expression combinée de MSH2 et MSH6 ou de MLH1 et PMS2 évoque d'avantage la présence d'une anomalie dans le gène *MSH2* et *MLH1*, respectivement. En effet, les protéines MMR fonctionnant sous forme d'hétérodimères MSH2-MSH6 et MLH1-PMS2, les protéines MSH6 et PMS2 sont déstabilisées en absence des protéines MSH2 et MLH1, respectivement (Richman *et al.*, 2015).

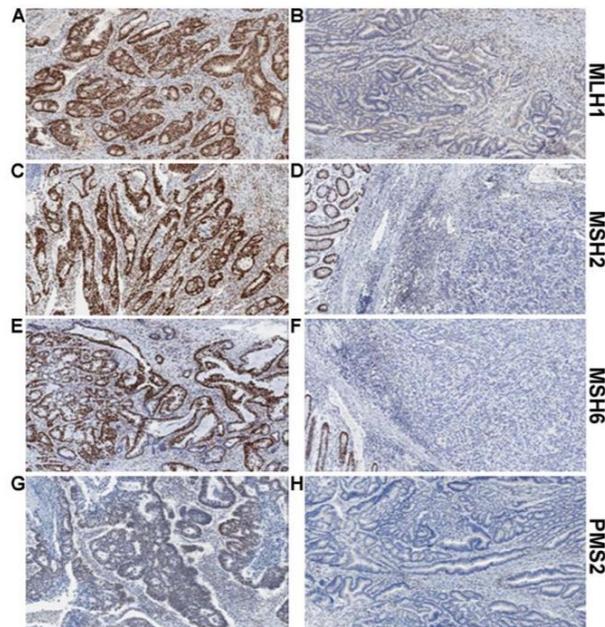


Figure 15. Analyse de l'expression tumorale des protéines MMR par immunohistochimie (d'après Richman *et al.*, 2015). Expression (A, C, E, G) ou absence d'expression (B, D, F, H) des protéines MLH1, MSH2, MSH6, PMS2, respectivement, détectée par immunohistochimie dans les tumeurs colorectales de patients évocateurs d'un syndrome de Lynch.

Les recommandations internationales : les critères d'Amsterdam et de Bethesda

L'ensemble de ces éléments, cliniques, familiaux et tumoraux, évocateurs d'un syndrome de Lynch a permis d'établir des recommandations internationales de bonnes pratiques standardisées et ce, dans le but de faciliter le diagnostic (Tableau 5 ; Giardiello *et al.*, 2014). C'est en 1991 que les premières recommandations ont été émises par le groupe collaboratif international de HNPCC sous la forme des critères d'Amsterdam I qui définissent que tout individu avec une histoire familiale de CCR présente un risque élevé de développer un CCR à un âge précoce, tout en excluant les syndromes de polyposes adénomateuses familiales. Ces recommandations ont fait l'objet d'une révision en 1999 et intègrent dès lors la présentation de certaines tumeurs extracoliques comme élément évocateur d'un syndrome de Lynch, constituant ainsi les critères d'Amsterdam II (Vasen *et al.*, 1999). En parallèle,

l'institut national du cancer américain a établi les critères de Bethesda, des recommandations s'appuyant notamment sur l'instabilité microsatellitaire détectée dans les tumeurs comme élément signature d'un diagnostic de syndrome de Lynch (Rodriguez-Bigas *et al.*, 1997). A l'instar des critères d'Amsterdam I, les critères de Bethesda ont également fait l'objet d'une révision quelques années plus tard, en 2004 (Umar *et al.*, 2004). Les critères de Bethesda révisés sont actuellement plus sensibles que ceux d'Amsterdam II et permettent ainsi l'identification de certaines formes cliniques atypiques du syndrome de Lynch qui ne répondent pas aux critères d'Amsterdam II.

Critères d'Amsterdam II	Critères de Bethesda révisés
<p>≥ 3 apparentés atteints d'un cancer du spectre étroit ou élargi avec (tous les critères sont nécessaires)</p> <hr/> <p>1 des parents unis aux 2 autres par un lien de parenté au 1^{er} degré</p> <hr/> <p>2 générations successives atteintes</p> <hr/> <p>≥ 1 parent atteint d'un cancer < 50 ans</p> <hr/> <p>Tumeurs prouvées histologiquement</p> <hr/> <p>Exclusion des polyposes adénomateuses familiales</p>	<p>Critères de Bethesda révisés (seul un critère suffit)</p> <hr/> <p>Cancer colorectal < 50 ans</p> <hr/> <p>2 cancers colorectaux synchrones ou métachrones ou 1 cancer colorectal associé à un autre cancer du spectre du syndrome de Lynch (quel que soit l'âge)</p> <hr/> <p>Cancer colorectal < 60 ans avec en histologie un adénocarcinome lieberkühnien avec instabilité des microsatellites (Statut MSI)</p> <hr/> <p>Cancer colorectal avec au moins 1 apparenté au 1^{er} degré présentant un cancer du spectre du syndrome de Lynch avant 50 ans</p> <hr/> <p>Cancer colorectal diagnostiqué avec ≥ 2 apparentés du 1^{er} ou 2^e degré ayant un antécédent de cancer du spectre du syndrome de Lynch (quel que soit l'âge)</p>

Tableau 5. Recommandations internationales de bonnes pratiques standardisées : les critères d'Amsterdam II et de Bethesda révisés (d'après Coffin *et al.*, 2019)

1.2.3.b. Le diagnostic moléculaire du syndrome de Lynch

Sur la base d'éléments évocateurs d'un syndrome de Lynch il peut être proposé au patient un diagnostic moléculaire afin de confirmer un cas de syndrome de Lynch (Figure 16). L'objectif est alors l'identification de l'anomalie constitutionnelle hétérozygote pathogène à l'origine de la déficience du système MMR. Les variations constitutionnelles à l'état homozygote dans les gènes *MMR* ne constituent pas une prédisposition au syndrome de Lynch mais sont associées au syndrome CMMRD (*Constitutional Mismatch Repair Deficiency*), un syndrome conduisant au développement simultané ou successif de différents types de tumeurs dès le plus jeune âge (Wang *et al.*, 1999a ; Buecher *et al.*, 2019). L'identification de la variation causale est indispensable à la bonne orientation du conseil génétique pour le patient et ses apparentés, d'un point de vue diagnostique et pronostique mais également thérapeutique.

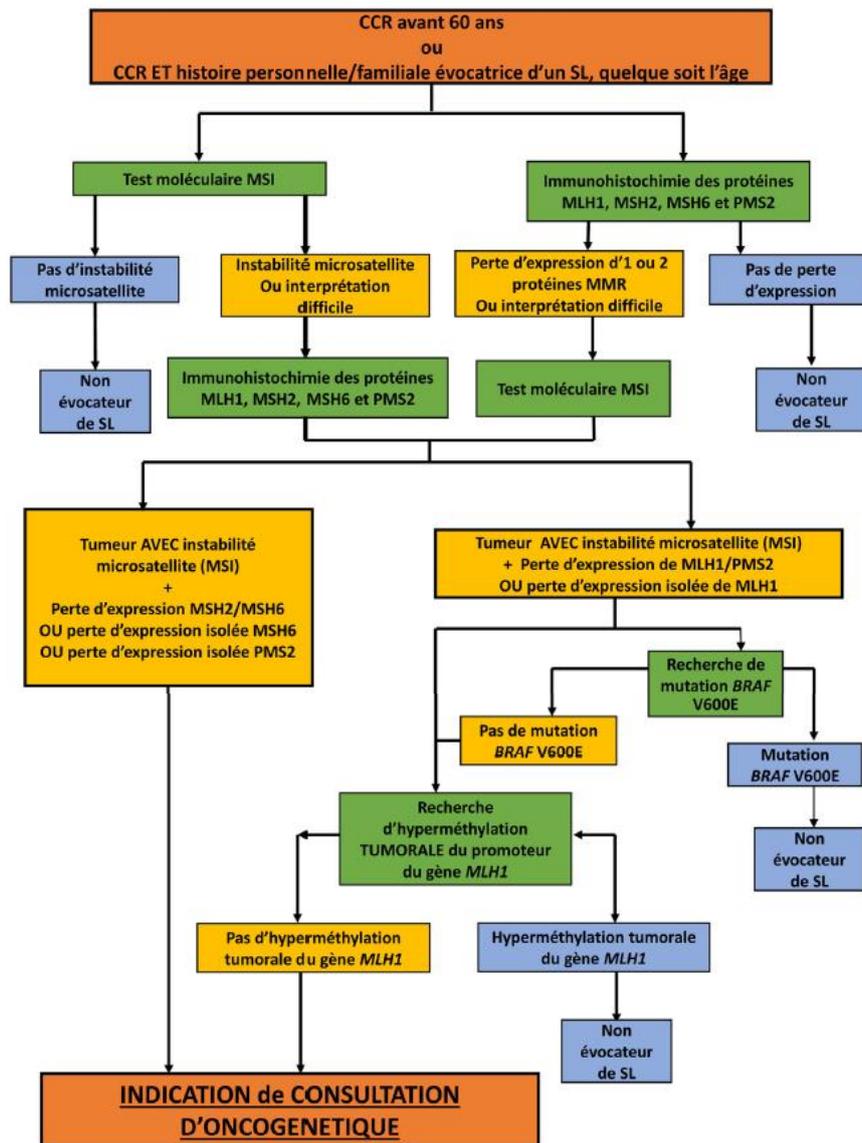


Figure 16. Recommandations au diagnostic moléculaire d'un syndrome de Lynch (d'après Coffin et al., 2019)

Au même titre que le diagnostic moléculaire du syndrome sein-ovaire, celui du syndrome de Lynch a été révolutionné par l'implémentation du NGS qui permet aujourd'hui d'analyser plusieurs gènes simultanément à moindre coût (Gallego et al., 2015, Baert-Desurmont et al., 2018). En France, le Groupe Génétique et Cancer a établi un panel de 14 gènes dont l'utilisation est recommandée pour toutes les formes héréditaires de CCR : *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*, *APC*, *MUTYH*, *BMPR1A*, *CDH1*, *POLD1*, *POLE*, *PTEN*, *SMAD4*, *STK11* (Dhooge et al., 2020). Cependant, il est important de noter que l'existence de plusieurs pseudogènes du gène *PMS2*, situés sur le même chromosome, complique la détection de variations dans ce gène (Nakagawa et al., 2004). Afin d'amplifier le gène *PMS2* et non l'un de ses pseudogènes, une technique de PCR long-range a spécifiquement été développée dans ce but (Vaughn et al., 2010).

Importance du diagnostic moléculaire du syndrome de Lynch

L'identification de l'altération génétique à l'origine de la maladie revêt un caractère essentiel d'un point de vue diagnostique, mais également pronostique, car elle va conditionner la prise en charge du cas index mais également le suivi médical de ses apparentés. Dans le cas d'individus non porteurs, cela permet de lever une surveillance inutile tandis que dans le cas d'individus porteurs de la variation pathogène, les cliniciens pourront proposer des examens de surveillance afin de dépister précocement tout développement tumoral. Des recommandations de prise en charge ont été émises par des consortia d'experts européens et américains telles qu'un individu porteur d'une variation pathogène dans l'un des gènes *MMR* se verra proposer des coloscopies annuelles complétées par des hystérosopies, des échographies utérines et des biopsies utérines pour les femmes (Vasen *et al.*, 2013). En France, un réseau national d'experts a été constitué afin d'établir des recommandations basées sur les données de la littérature scientifique ainsi que sur les différentes recommandations émises à l'international (Figure 17 ; Coffin *et al.*, 2019).

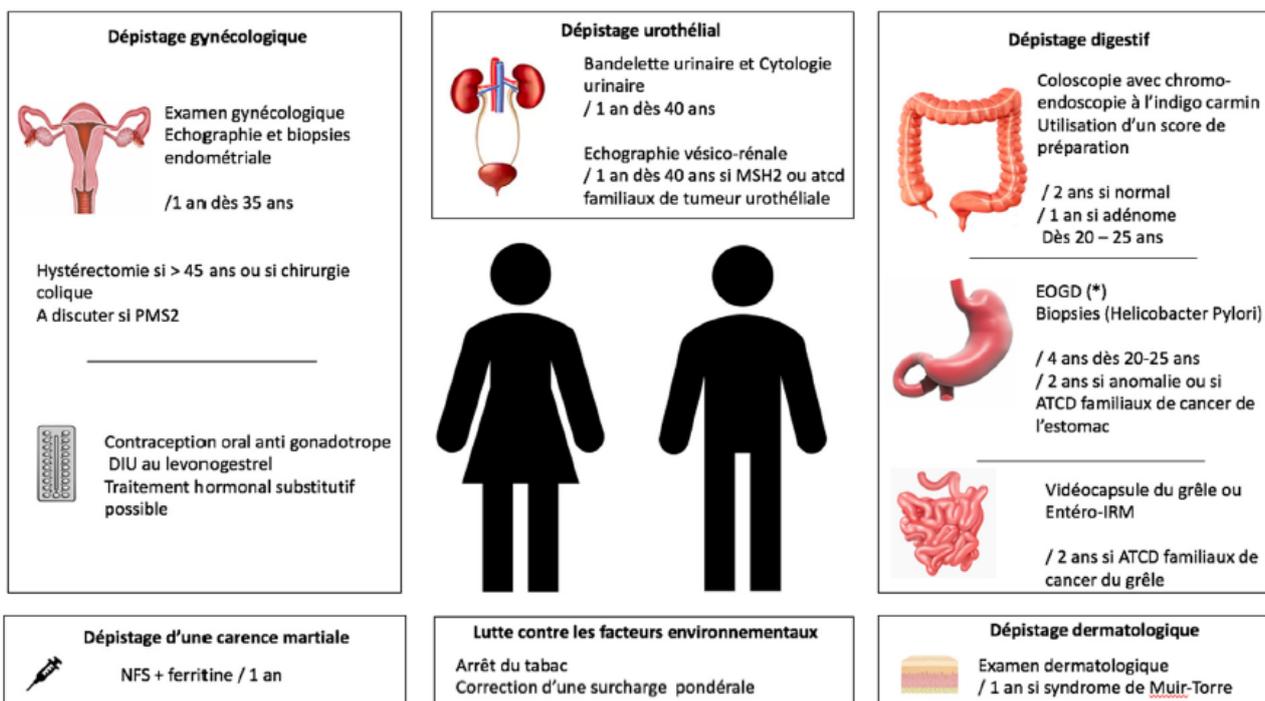


Figure 17. Recommandations proposées par le réseau français pour la surveillance d'un individu porteur d'une altération pathogène d'un des gènes *MMR* (d'après Coffin *et al.*, 2019)

Ainsi, un individu porteur d'une variation pathogène d'un des gènes *MMR* se verra proposer un dépistage digestif consistant en une coloscopie tous les deux ans dès l'âge de 20/25 ans pouvant être rapproché à un intervalle d'un an si un polype adénomateux ou festonné a été mis en évidence lors du précédent examen. Une surveillance duodénale est également recommandée par endoscopie

digestive haute tous les quatre ans dès 20/25 ans ou tous les deux ans dans le cas d'antécédents familiaux de cancer de l'estomac. A partir de 40 ans, un dépistage urothélial est conseillé tous les ans avec une cytologie urinaire ainsi qu'une échographie vésico-rénale dans le cas d'individus porteurs d'une altération pathogène du gène *MSH2* ou d'antécédents familiaux de tumeurs urothéliales. Pour les femmes, un examen gynécologique, une échographie et des biopsies endométriales sont recommandés tous les ans dès 35 ans. Une hystérectomie et une salpingo-oophorectomie prophylactiques peuvent également être proposées après 45 ans afin de diminuer le risque d'atteinte d'un cancer de l'endomètre ou des ovaires. Enfin, il sera conseillé aux individus porteurs d'éviter ou d'arrêter le tabac et également de surveiller leur charge pondérale. De façon tout à fait remarquable, des études prospectives ont permis de démontrer qu'un suivi régulier par coloscopie des individus à risque permettait de réduire significativement l'incidence et la mortalité des CCR (Yurgelun and Hampel, 2018). Néanmoins, il n'a pas été démontré qu'un suivi gynécologique quel qu'il soit, ait un réel impact sur la détection précoce des cancers de l'endomètre ou des ovaires associés à un syndrome de Lynch (Yurgelun and Hampel, 2018).

Intérêt thérapeutique du diagnostic moléculaire du syndrome de Lynch

Dans le cas d'individus déjà atteints d'un cancer, l'identification de l'altération constitutionnelle revêt un caractère essentiel d'un point de vue thérapeutique. En particulier, il a été rapporté que l'utilisation de certains traitements anti-cancéreux classiques, tels que les sels de platines et le 5-fluorouracile, étaient inefficaces contre les tumeurs déficientes en système MMR (Fink *et al.*, 1998 ; Sargent *et al.*, 2010). Il est donc recommandé de traiter ces tumeurs avec d'autres agents anti-cancéreux ou d'ajouter des adjuvants capables de rétablir le bénéfice de la chimiothérapie. De façon importante, l'identification de l'altération pathogène dans un des gènes *MMR* conditionne aujourd'hui l'accès à certaines thérapies ciblées. Il a été mis en évidence que les tumeurs associées au syndrome de Lynch pouvaient être traitées par des approches thérapeutiques basées sur le concept de létalité synthétique (Martin *et al.*, 2010 ; Guillotin and Martin, 2014). Au même titre que les thérapies ciblées utilisées dans un contexte de syndrome sein-ovaire, ces thérapies sont basées sur la perte de fonction concomitante de deux voies de réparation complémentaires : l'inhibition des ADN polymérases impliquées dans la voie de réparation BER conduit à une accumulation des lésions oxydatives au niveau de l'ADN des cellules tumorales qui, déficientes en activité MMR, ne peuvent réparer ces dommages. Cette thérapie entraîne donc une surcharge de lésions non réparables dans les cellules déficientes en système MMR conduisant ainsi à la mort cellulaire. Il est toutefois important de noter que la réponse thérapeutique peut varier selon le gène altéré. Par exemple, les cellules tumorales déficientes en protéines *MSH2* sont particulièrement sensibles au méthotrexate, un agent induisant des dommages

oxydatifs, alors que les cellules déficientes en protéines MLH1 sont résistantes à cette drogue (Martin *et al.*, 2009).

Dans le cas de tumeurs à MSI, l'immunothérapie a été reconnue comme un traitement pertinent (Valle *et al.*, 2019). En effet, l'accumulation de variations de type *frameshift* engendrées par l'instabilité microsatellitaire de certaines séquences répétées au sein de régions codantes conduit à la production de néoantigènes reconnus par les cellules lymphocytaires T CD8+ (Maby *et al.*, 2016). Les tumeurs à MSI sont donc caractérisées par une densité élevée de lymphocytes infiltrant la tumeur (TIL, *Tumor Infiltrating Lymphocytes*) (Smyrk *et al.*, 2001). Ainsi, il pourrait être attendu que ces tumeurs soient éliminées directement par le système immunitaire. Ce n'est cependant pas le cas car une grande quantité de molécules immunosuppressives (PD-1, PD-L1, CTLA-4, LAG-3 et IDO) est exprimée par les cellules tumorales et leur microenvironnement immunitaire (Llosa *et al.*, 2015). Récemment, il a été montré dans 3 études de phase II que l'administration d'anticorps monoclonaux anti PD-1 (pembrolizumab, nivolumab) seuls (Le *et al.*, 2015 ; Overman *et al.*, 2017) ou en combinaison avec un anti-CTLA-4 (ipilumab) (Overman *et al.*, 2018) permettait d'assurer un contrôle tumoral prolongé chez des patients qui ne répondent pas aux traitements conventionnels. Le bénéfice clinique observé pour ces patients a ainsi permis à deux anticorps monoclonaux, le pembrolizumab et le nivolumab, d'être autorisés par la FDA (*Food and Drug Administration*) aux Etats-Unis comme traitement des tumeurs à MSI. Cependant, en Europe, ces molécules ne sont pas disponibles pour cette indication et les patients doivent être orientés vers des essais thérapeutiques pour pouvoir bénéficier de cette immunothérapie. Un essai de phase III international initié par le professeur Thierry André de l'hôpital AP-HP Saint-Antoine à Paris a permis de démontrer l'impact significatif du pembrolizumab sur la survie sans progression tumoral de patients atteints d'un cancer colorectal métastatique à MSI (André *et al.*, 2020). Ces résultats valident l'intérêt de l'immunothérapie en traitement de première intention dans le cas de CCR métastatiques à MSI et devraient permettre l'obtention d'une autorisation européenne de mise sur le marché.

II. La problématique de l'interprétation clinique des variations génétiques

La grande variabilité inter-individuelle du génome humain

La plus grande réussite scientifique du XX^{ème} siècle fut, sans conteste, la découverte de la structure en double hélice de l'ADN par Rosalind Franklin, Maurice Wilkins, James Watson et Francis Crick (Watson and Crick, 1953). Cette révolution scientifique a marqué un tournant décisif pour la biologie moléculaire et la génétique humaine et a entraîné l'essor de multiples inventions parmi lesquelles, le séquençage de la molécule d'ADN. La première technologie de séquençage de l'ADN a été développée par le biochimiste anglais Frederick Sanger et porte désormais son nom (Sanger *et al.*, 1977). En 1980, le développement du séquençage Sanger a rendu envisageable une idée qui faisait rêver plus d'un biologiste : cartographier le génome humain entier. Le programme *Human Genome Project* a alors vu le jour grâce à un effort collaboratif international sans précédent regroupant les seize plus grands instituts de biotechnologie du monde. Il aura fallu treize ans et près de 3 milliards de dollars pour enfin parvenir à déchiffrer les 3,4 milliards de bases du génome humain, constituant ainsi une avancée unique et primordiale pour la génétique clinique (International Human Genome Sequencing Consortium, 2004 ; Shendure *et al.*, 2017).

Presque vingt ans plus tard, le séquençage du génome entier d'un individu peut désormais être réalisé pour moins de 1 000 dollars et en moins de vingt-quatre heures (*National Human Genome Research Institute*, <https://www.genome.gov>). Ceci a été rendu possible grâce au développement de nouvelles technologies de séquençage à haut débit (NGS, *Next-Generation Sequencing*). Au début des années 2000, fort des résultats du programme *Human Genome Project*, les premiers séquenceurs à haut débit, dits de première génération, ont été rapidement développés afin d'optimiser le coût et le temps de séquençage d'un génome individuel (Shendure *et al.*, 2017). Depuis, les technologies de séquençage à haut débit n'ont cessé d'évoluer, faisant place aux séquenceurs de deuxième, troisième puis quatrième génération, se renouvelant sans cesse pour séquencer toujours plus vite et toujours moins cher.

L'essor de ces technologies de séquençage à haut débit a été mis à profit par de très importants consortiums internationaux dans le but d'appréhender la diversité du génome humain. Ainsi, le projet 1000 Génomes a permis de mettre en évidence le très large spectre de la variabilité génétique du génome humain (The 1000 Genomes Project Consortium, 2015). Sur la base du séquençage du génome de 2 504 individus provenant de 26 populations différentes, plus de 88 millions de variations

différentes ont été détectées. Ces variations peuvent être regroupées en 3 catégories : (i) les variations ponctuelles d'une base (SNV, *Single Nucleotide Variant*) les plus fréquentes (84,7 millions), (ii) les petites insertions/délétions (indels) (3,6 millions) mais aussi (iii) les réarrangements de grande taille ou variations structurales (60 000) (The 1000 Genomes Project Consortium, 2015). La comparaison de la séquence du génome d'un individu donné avec celle du génome de référence a alors fait apparaître en moyenne 4 à 5 millions de variations ponctuelles (Tableau 6 ; Eichler *et al.*, 2019).

Class	Size of Variant <i>bp</i>	No. per Genome [†]	Size of Region Affected <i>Mbp</i>	Percent of Genome
Single-nucleotide variants	1	4,000,000–5,000,000	4–5	0.078
Insertions–deletions	1–49	700,000–800,000	3–5	0.069
Structural variants	>50	23,000–28,000	10–12	0.19
Inversions	>50	153 [‡]	23 [‡]	0.397
Multi-copy-number variants [§]	>1000	Approximately 500	12–15	0.232

Tableau 6. Variabilité interindividuelle du génome (d'après Eichler *et al.*, 2019)

De façon importante, cette très grande variabilité génétique ne se restreint pas uniquement aux séquences non-codantes du génome humain mais elle impacte également les régions codant les protéines. En 2016, les données issues du consortium international ExAC (*Exome Aggregation Consortium*) à partir du séquençage d'exomes de 60 706 individus d'origines diverses constituent pour la première fois un catalogue exhaustif des variations génétiques identifiées dans les séquences codantes humaines (Lek *et al.*, 2016). Ces travaux ont ainsi permis d'identifier 7,4 millions de variations (en moyenne une toutes les 8 bases) dont 72% correspondaient à des variations nouvelles (*i.e.* non répertoriées dans les autres bases de données) et 54% à des singletons (*i.e.* variations portées par un unique individu). Ce répertoire de variations génétiques humaines ne cesse de s'étendre du fait de l'agrégation de nouvelles données de séquençage. Aujourd'hui, l'ensemble de ces données recouvre l'analyse non seulement des SNV mais également des variants structuraux pour un nombre total de 141 456 individus (*i.e.* 125 748 exomes et 15 708 génomes) (Karczewski *et al.*, 2020). Ces informations sont désormais répertoriées dans la base de données gnomAD (*the genome Aggregation Database*, <https://gnomad.broadinstitute.org/>). Régulièrement mise à jour et libre d'accès, cette base constitue une ressource primordiale dans le domaine de la génétique humaine et tout particulièrement du fait de son utilisation pour l'interprétation clinique des variations identifiées dans le génome des patients. Il est désormais estimé que l'exome de chaque individu présente, en moyenne, 21 000 SNV en comparaison avec celui de référence (données internes de notre laboratoire Inserm U1245, Rouen).

La très grande majorité de ces variations correspond à des polymorphismes très fréquents dans la population générale (MAF, *Minor Allele Frequency* $\geq 1\%$). Toutefois, chaque individu est porteur d'environ 1 400 variations rares (MAF $< 1\%$) voire uniques. Cette très grande variabilité inter-individuelle du génome humain a de très fortes implications dans la recherche de corrélations génotype/phénotype dans le contexte des maladies à déterminisme mendélien et donc, dans l'interprétation biologique et clinique de la pléthore de variations identifiées dans le génome des patients.

La variabilité génétique appréhendée dans un contexte de maladie mendélienne

Grâce au développement et à l'implémentation du séquençage à haut débit de l'ADN en routine, la génétique médicale humaine est entrée dans une nouvelle ère. En effet, à la suite de la réduction importante des coûts ainsi que la miniaturisation des séquenceurs, le NGS est devenu une technologie à la portée d'un grand nombre de laboratoires de recherche mais aussi de diagnostic moléculaire rendant ainsi beaucoup plus aisé le criblage des mutations dans le génome de patients. Dans un contexte de maladie mendélienne, l'enjeu majeur n'est donc plus la détection des variations dans le génome des patients mais bel et bien, leur interprétation biologique et clinique, c'est-à-dire la détermination de leur caractère neutre ou pathogène. De façon importante, l'identification de la mutation causale, à l'origine de la maladie, aura de réelles conséquences pour la prise en charge des patients et des apparentés. Ainsi, l'approche traditionnelle de la médecine a fortement évolué et s'oriente de plus en plus vers une médecine génomique de précision, dépendante de l'identification de la variation pathogène.

L'interprétation clinique des variations génétiques identifiées chez un patient évocateur d'une maladie à déterminisme mendélien est bien sûr considérablement facilitée par la connaissance en amont des gènes impliqués dans la pathologie. Toutefois, la problématique de l'interprétation reste importante si les gènes mis en cause présentent un très large spectre mutationnel, c'est-à-dire si un très grand nombre de variations différentes et plus ou moins rares sont détectées dans ces gènes. Avec l'avènement du NGS en diagnostic moléculaire, la quantité de variations différentes identifiées dans le génome des patients s'est encore accrue, rendant alors encore plus prégnant le défi de l'interprétation des variations génétiques (Vears *et al.*, 2017). De plus, les enjeux associés à cette problématique se sont considérablement étendus avec le développement de certaines thérapies ciblées (Evans *et al.*, 2015 ; Rehm *et al.*, 2015). Ainsi, dans certains cas, l'absence d'interprétation des variations identifiées chez les patients constitue un obstacle majeur à la médecine génomique de précision. Pour l'ensemble de ces raisons, l'établissement d'un cadrage précis pour l'interprétation clinique des variations avec la

définition de critères rigoureux et spécifiques s'est imposé comme une nécessité afin d'assurer des pratiques harmonisées de classification des variations entre les différents laboratoires de diagnostic moléculaire.

II.1. Les recommandations générales dans un contexte de maladies mendéliennes

En l'absence de méthodes et de règles prédéfinies pour l'interprétation clinique des variations génétiques, certaines inconsistances ou discordances de classification pouvaient être observées entre laboratoires de diagnostic moléculaire, avec des conséquences potentiellement importantes pour la prise en charge médicale des patients et de leurs apparentés. Ainsi, en 2015, des analyses de la classification des variations répertoriées dans la base de données ClinVar ont mis en lumière que, pour 12 895 variations associées à deux ou plusieurs entrées de soumission, l'interprétation différait dans 17% des cas, c'est-à-dire pour 2 229 variations (Rehm *et al.*, 2015). Devant le besoin urgent d'harmoniser les pratiques utilisées pour l'interprétation des variations dans les laboratoires de diagnostic, le consortium représenté par l'*American College of Medical Genetics and Genomics* (ACMG) et l'*Association for Molecular Pathology* (AMP) a émis, dès 2000, des recommandations pour l'établissement d'une méthode de classification (Kazazian *et al.*, 2000). Les propositions initiales de cadrage ont été par la suite améliorées notamment en définissant rigoureusement les critères pris en compte ainsi qu'en mettant en place une pondération adaptée des différents éléments de preuve en faveur du caractère pathogène ou neutre des variations (Richards *et al.*, 2008, 2015).

Le groupe de travail ACMG-AMP a adopté le système d'interprétation des variations basé sur cinq classes initialement établies par le Centre international de Recherche sur le Cancer (IARC, *International Agency for Research on Cancer*) (Plon *et al.*, 2008), à savoir : (i) les variations neutres ou classe 1, (ii) les variations probablement neutres ou classe 2, (iii) les variations de signification inconnue (VSI) ou classe 3, (iv) les variations probablement pathogènes ou classe 4 et (v) les variations pathogènes ou classe 5. Chacune de ces classes est ainsi associée à des recommandations spécifiques de prise en charge du patient porteur et de ses apparentés, adaptées en fonction des pathologies. Les porteurs de variations de classes 1 et 2, s'agissant de variations sans impact causal, ne présenteront pas plus de suivi spécifique que la population générale. Les variations de classe 3 ne pouvant être interprétées dans l'état actuel des connaissances ne seront donc pas utilisées dans le cadre du conseil génétique car il est impossible de les relier à la pathologie sans toutefois les exclure. Quant aux variations de classes 4 et 5, elles sont directement associées à l'étiologie de la maladie et de ce fait les individus porteurs se verront proposer une prise en charge spécifique. L'orientation de la classification d'une variation dans une de ces cinq catégories repose sur un faisceau d'arguments qui, pris

individuellement, présentent rarement un poids suffisant à l'interprétation de la variation, mais qui, par agrégation, permettent dans certains cas une classification précise (Richards *et al.*, 2015).

Prise en considération de la nature de la variation

Certaines variations sont fortement soupçonnées d'être pathogènes étant donné leur nature. Il s'agit (i) des remaniements de grande taille (délétion, duplication, inversion d'un ou plusieurs exons, par exemple), (ii) des variations non-sens, (iii) des petites insertions/délétions qui décalent le cadre de lecture (*frameshift*) et (iv) des variations localisées au niveau des positions les plus conservées des sites canoniques d'épissage ($IVS\pm 1/2$) (Tableau 7 ; Richards *et al.*, 2015). En effet, ces variations conduisent souvent à l'introduction d'un codon stop prématuré (PTC, *Premature Stop Codon*) au niveau de l'ARNm qui peut alors être ciblé pour dégradation par un système de surveillance cellulaire connu sous le nom de *nonsense-mediated mRNA decay* (NMD) (Popp and Maquat, 2013 ; Kurosaki and Maquat, 2016). Si les transcrits porteurs du PTC échappent au NMD, ils sont susceptibles d'être traduits et de générer des protéines tronquées. Dans ces deux situations, la conséquence la plus souvent attendue est l'absence totale de protéine fonctionnelle. Sur la base de ces arguments, ces variations sont généralement considérées comme des mutations à l'origine d'une perte de fonction et sont donc associées au critère le plus fort en faveur de la pathogénicité (PVS1, *Pathogenic Very Strong criterion*) (Richards *et al.*, 2015).

Pour les autres types de variations (faux-sens, synonymes, indels en phase, introniques en dehors des positions $IVS\pm 1/2$), l'interprétation biologique et clinique est plus difficile et requiert des analyses complémentaires. A défaut d'éléments additionnels, beaucoup de ces variations sont aujourd'hui classées VSI, ce qui constitue un obstacle majeur à l'optimisation de la prise en charge médicale des patients et de leurs apparentés.

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population Data	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
Functional Data	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

Tableau 7. Recommandations de l'ACMG-AMP pour la classification des variations (d'après Richards *et al.*, 2015)

Utilisation de la fréquence allélique de la variation dans la population générale

Dans un premier temps, il est essentiel de collecter des informations relatives à la fréquence allélique de la variation dans la population générale et d'effectuer des analyses comparatives sur des cohortes d'individus témoins et de patients (études cas-témoin). L'accès aux informations relatives à la fréquence allélique de la variation est permis par l'interrogation de bases de données populationnelles qui agrègent l'ensemble des variations découvertes dans le génome d'individus appartenant à différentes populations (Tableau 8 ; Richards *et al.*, 2015). Plusieurs projets de séquençage à grande échelle ont été menés et ont permis la constitution de ces bases de données telles que dbSNP (*Single Nucleotide Polymorphism database*) (Sherry *et al.*, 2001), première grande base de données répertoriant les variations génomiques identifiées en population, le projet 1000Genomes qui comprend les données de séquençage de plus de 2 500 génomes d'individus issus de 26 populations (Sudmant *et al.*, 2015), le projet ESP (*Exome Sequencing Project*) qui rassemble les exomes de près de 7 000 individus américains, le consortium ExAC cumulant les données de

séquençage d'exomes de plus de 60 000 individus issus de 6 groupes de populations, et enfin, plus récemment la base de données gnomAD (Lek *et al.*, 2016) qui permet aujourd'hui l'accès aux données de plus de 125 000 exomes et 15 000 génomes (Karczewski *et al.*, 2020). Il est à noter que les informations issues de ces bases de données doivent être utilisées avec prudence dans certains cas. En effet, certaines d'entre elles hébergent des données de séquençage issues de patients ayant développé des maladies dont l'étiologie peut être d'origine génétique, comme des cancers, des maladies neurodégénératives ou des pathologies cardio-vasculaires. De plus, si les individus dont le génome a été séquençé ne présentent pas de pathologie au moment du séquençage, il n'est pas impossible qu'une partie d'entre eux développent une maladie mendélienne ultérieurement, faussant ainsi l'interprétation des variations dont ils sont porteurs. De plus, même si ces bases de données sont issues d'efforts internationaux, elles ne sont pas représentatives de toutes les sous-populations.

Population Databases	
Exome Aggregation Consortium http://exac.broadinstitute.org/	Database of variants found during exome sequencing of 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects as well as related individuals were excluded.
Exome Sequencing Project http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of several large cohorts of individuals of European and African American ancestry. Includes coverage data to inform the absence of variation.
1000 Genomes http://browser.1000genomes.org	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity compared to EVS but also contains lower quality data and some cohorts contain related individuals.
dbSNP http://www.ncbi.nlm.nih.gov/snp	Database of short genetic variations (typically 50 bp or less) submitted from many sources. May lack details of originating study and may contain pathogenic variants.
dbVar http://www.ncbi.nlm.nih.gov/dbvar	Database of structural variation (typically greater than 50 bp) submitted from many sources.
Disease Databases	
ClinVar http://www.ncbi.nlm.nih.gov/clinvar	Database of assertions about the clinical significance and phenotype relationship of human variation.
OMIM http://www.omim.org	Database of human genes and genetic conditions that also contains a representative sampling of disease-associated genetic variants.
Human Gene Mutation Database http://www.hgmd.org	Database of variant annotations published in the literature. Requires fee-based subscription for much of the content.
Locus/Disease/Ethnic/Other-Specific Databases http://www.hgvs.org/dblist/dblist.html http://www.lovd.nl	The HGVS site developed a list of thousands of different databases that provide variant annotations on specific subsets of human variation. A large percentage of databases are built in the LOVD system.
DECIPHER http://decipher.sanger.ac.uk	A molecular cytogenetic database for clinicians and researchers linking genomic microarray data with phenotype using the Ensembl genome browser.

Tableau 8. Bases de données de population générale et bases de données mutationnelles spécifiques des maladies mendéliennes (d'après Richards *et al.*, 2015)

Une variation susceptible d'être pathogène doit présenter une fréquence allélique compatible avec la prévalence de la maladie dans une population donnée (Figure 18 ; Manolio *et al.*, 2009). Dans l'absolu, ce critère doit également prendre en compte le spectre mutationnel du ou des gènes impliqués ainsi que le mode de transmission dominant ou récessif de la maladie. Les maladies mendéliennes étant considérées comme rares, le seuil recommandé par l'ACMG de façon très conservative est celui de 5% (Richards *et al.*, 2015) mais il est souvent ajusté en fonction de la maladie à un seuil inférieur ou égal à 1%. En pratique, ce critère est utilisé pour filtrer les variations présentant une MAF supérieure à ce seuil qui sont alors qualifiées de polymorphismes et classées comme neutres sur la seule base de cette information (BS1, *Benign Strong criterion*) (Tableau 7 ; Richards *et al.*, 2015). Il est considéré par ailleurs que l'absence totale de la variation dans les bases de données populationnelles constitue un argument de poids modéré (PM2, *Pathogenic Moderate criterion*) en faveur de la pathogénicité (Tableau 7 ; Richards *et al.*, 2015).

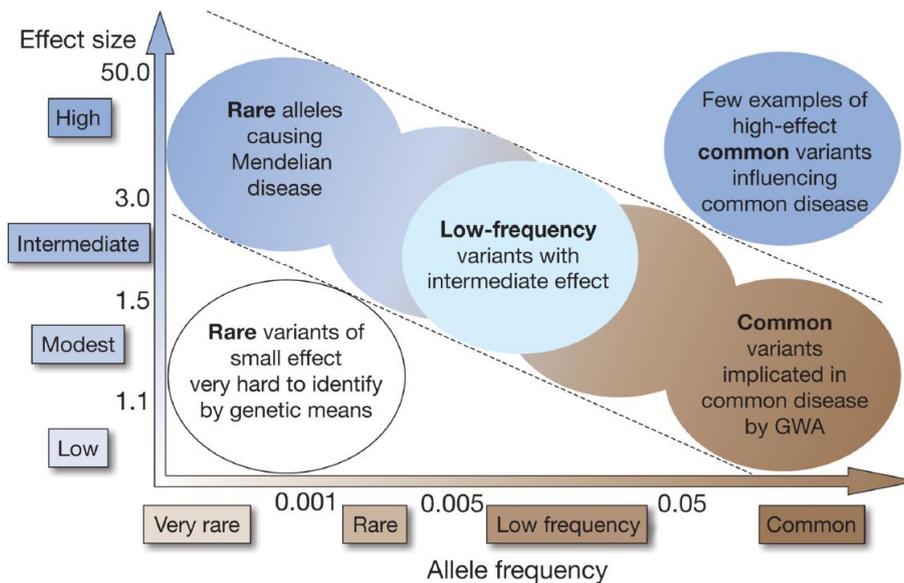


Figure 18. Effet génétique des variations en fonction de leur fréquence allélique (d'après Manolio *et al.*, 2009). Les variations présentant des associations indiquées entre les lignes pointillées en diagonale sont celles revêtant le plus d'intérêt en génétique médicale.

Utilisation des bases de données établies à partir de cohortes de patients

Dans un contexte de génétique médicale, des efforts nationaux et internationaux majeurs ont été entrepris afin de développer des bases de données mutationnelles en lien direct avec les maladies à déterminisme mendélien (Tableau 8). L'objectif est alors de compiler toutes les informations relatives aux variations de séquence détectées dans les gènes connus pour être impliqués dans une maladie mendélienne et identifiées chez les individus ayant développé cette pathologie. La puissance de ces ressources réside dans la corrélation établie entre le phénotype et le génotype. Ces bases de

données sont largement et régulièrement implémentées grâce aux informations soumises principalement par les laboratoires de diagnostic moléculaire. Une des ressources de référence dans ce domaine est la base ClinVar (Landrum *et al.*, 2014, 2018) qui répertorie à ce jour plus de 840 000 variations identifiées chez des patients. Outre le nombre considérable et croissant de variations enregistrées, l'un des principaux avantages de cette base internationale réside dans l'indication d'une interprétation clinique. En effet, non seulement le phénotype associé à la variation est rapporté mais, dans la plupart des cas, une classification est également proposée par le laboratoire à l'origine de la soumission. Il est important de noter que la fiabilité des interprétations proposées est présentée sous la forme de l'attribution d'un nombre plus ou moins important d'étoiles, en fonction du nombre et de la robustesse des arguments soumis. Une autre base de données importante de corrélation génotype/phénotype dans un contexte de maladie à déterminisme mendélien est celle de HGMD (*Human Gene Mutation Database*) (Stenson *et al.*, 2017). Cette ressource répertorie aujourd'hui plus de 270 000 variations dans plus de 11 000 gènes et indique toutes les informations associées sur la base de la collecte de données bibliographiques. Ces informations n'étant pas toujours exhaustives, il est recommandé de consulter indépendamment la littérature médicale et scientifique associée aux variations. En parallèle de ces ressources, se sont développées des plateformes spécifiques de locus telle que la base de données LOVD (*Leiden Open Variation Database*) (Fokkema *et al.*, 2011) qui intègre en plus des données transmises par les laboratoires de diagnostic des soumissions issues des laboratoires de recherche ou collectées par les curateurs. En France, une initiative importante a été la création de bases de données, également spécifiques de locus, supportées par la plateforme UMD (*Universal Mutation Databases*) (Bérout *et al.*, 2000). Les variations identifiées par les laboratoires de diagnostic moléculaire sont répertoriées et font l'objet d'une interprétation rigoureuse sur la base d'une expertise développée au sein d'un réseau national de groupes de travail spécialisés de chaque maladie.

Il est important de noter ici que la majorité de ces bases de données sont accessibles librement en ligne, à l'exception de HGMD qui requiert une licence pour accéder à l'ensemble des données. Elles constituent donc des outils essentiels de partage des connaissances dans le domaine de la génétique médicale, à l'échelle nationale et internationale. La consultation des bases de données de variations génétiques identifiées chez les patients permet ainsi aux laboratoires de diagnostic moléculaire de déterminer si la variation d'intérêt a déjà été rapportée et si oui, d'avoir accès aux informations associées, notamment à celles relatives à sa classification.

L'évaluation du caractère neutre ou pathogène d'une variation s'appuie également sur l'analyse combinée de multiples données cliniques, familiales, bioinformatiques et biologiques.

Mis à profit des données cliniques et familiales

Le fait qu'un patient présente un phénotype correspondant au spectre connu des caractéristiques cliniques de l'altération d'un gène ne constitue pas une preuve seule de pathogénicité pour la variation d'intérêt localisée dans ce gène. En effet, presque tous les patients dont le génome fait l'objet d'un séquençage ciblé en quête d'une anomalie présente le phénotype associé à la maladie, c'est d'ailleurs un des critères d'orientation vers une consultation de conseil génétique.

Au sein des familles de patients, il convient de vérifier la co-ségrégation de la variation d'intérêt avec la maladie (Thompson *et al.*, 2003 ; Jarvik and Browning, 2016). En effet, une variation identifiée spécifiquement chez tous les individus atteints et absente chez les individus sains constituera un argument fort en faveur du caractère pathogène de la variation (Tableau 7 ; Richards *et al.*, 2015). Cet argument est particulièrement puissant si la co-ségrégation du génotype avec le phénotype est observée pour un grand nombre d'apparentés éloignés dans l'arbre généalogique et si l'analyse est conduite pour une même variation dans plusieurs familles non apparentées. En revanche, des familles de petite taille ou des données incomplètes rendront impossible cette analyse. De plus, la rareté de certaines variations, parfois identifiées dans une seule famille avec peu d'apparentés testés, rend impossible ce type d'analyse. Dans le cas d'un patient sans aucune histoire familiale, il est également possible que la variation soit survenue *de novo*, c'est-à-dire due à un accident génétique survenu au niveau de l'un des gamètes des parents de l'individu ou au stade post-zygotique. La présence d'une variation de cette nature est susceptible de constituer un argument en faveur de son implication dans la maladie (*Pathogenic Moderate/Strong criterion*) (Richards *et al.*, 2015). Il est également à noter que les données de ségrégation sont plus difficiles à interpréter dans le cas de gènes associés à une pénétrance incomplète ou dans le cas de la possible présence de phénocopies.

Informations apportées par les données de co-occurrence de variations

Pour un patient donné, plusieurs variations peuvent être détectées au niveau d'un même gène impliqué dans la maladie. Lorsque deux variations sont identifiées en *trans* dans un gène associé à une maladie autosomique dominante et que l'une des deux est classée pathogène sans ambiguïté, la seconde pourra éventuellement être classée neutre sur la base de cette observation. Cet argument est en particulier recevable si, pour une pathologie donnée, les individus porteurs à l'état hétérozygote composite de deux variations pathogènes dans ce gène ne sont pas supposés viables ou expriment un phénotype très différent de celui décrit pour la maladie avec un mode de transmission autosomique dominante. Il est à noter que déterminer la phase de deux variations (en *trans* ou *cis*) n'est pas toujours aisé, en particulier si l'analyse génétique des parents n'est pas possible et si les variations sont

distantes. Le développement des techniques de séquençage de 3^{ème} génération rendant possible l'analyse de très longs fragments devrait permettre, dans un futur proche, de pallier cette dernière limite.

Intégration des prédictions *in silico*

En parallèle des informations issues des bases de données, de la littérature et des analyses cliniques et familiales, il convient d'explorer l'impact biologique potentiel de la variation d'intérêt. De nombreux algorithmes de prédiction *in silico* ont ainsi été développés ou adaptés dans le but de modéliser l'effet d'une variation sur le processus d'épissage de l'ARN (l'ensemble de ces outils sont explicités de manière exhaustive dans la partie IV.2.) et sur la fonction ou la structure de la protéine.

De nombreux outils bioinformatiques ont été développés afin de prédire l'impact fonctionnel des variations faux-sens (Tableau 9) comme, par exemple, les algorithmes SIFT (*Sorting Intolerant From Tolerant* ; <http://blocks.fhcrc.org/sift/SIFT.html>) (Ng and Henikoff, 2001), Align-GVGD (http://agvgd.iarc.fr/agvgd_input.php) (Tavtigian *et al.*, 2008), PolyPhen-2 (*Polymorphism Phenotyping v2*, <http://genetics.bwh.harvard.edu/pph2>) (Adzhubei *et al.*, 2013), MutationTaster (Schwarz *et al.*, 2014) ou CADD (*Combined Annotation Dependent Depletion*) (Kircher *et al.*, 2014). Ces approches établissent, en général, un score de pathogénicité qui peut être attribué sur la base de différents critères tels que la conservation phylogénétique de l'acide aminé remplacé, les différences de propriétés physico-chimiques entre les acides aminés en question ou l'impact prédit de ce changement sur la structure de la protéine. Selon l'algorithme consulté, ces différents éléments peuvent être pris en compte seuls ou en analyse combinatoire. Parmi les outils de prédiction *in silico* les plus utilisés, on peut tout d'abord distinguer le logiciel SIFT qui établit un score en fonction de l'homologie de séquence protéique et considère que plus un acide aminé est conservé et plus les propriétés physico-chimiques de l'acide aminé muté sont différentes de l'acide aminé sauvage, plus la variation aura une probabilité élevée d'être pathogène. L'algorithme PolyPhen-2, quant à lui, repose sur une étude de l'homologie de séquence mais combine cette analyse à un autre paramètre, celui de la structure de la protéine afin de déterminer si la substitution conduit potentiellement à une modification du repliement et de la structure tridimensionnelle de la protéine. Enfin, le logiciel MutationTaster2 s'appuie non seulement sur la conservation protéique au cours de l'évolution mais intègre également des données de fréquence extraites des bases de données de populations générales ainsi que les informations de pathogénicité reportées dans les bases de données de variations génétiques identifiées chez des patients.

Tool	Category	Coding (missense only)	Indel	Non-coding	Method summary
SIFT ¹¹²	Missense prediction	Y (Y)	N	N	The degree of protein sequence conservation is used to predict the impact of a missense variant
PolyPhen2 (REF. 113)	Missense prediction	Y (Y)	N	N	Uses protein sequence and structure to predict the impact of a missense variant
FATHMM ¹¹⁴	Missense prediction	Y (Y)	N	N	Uses protein sequence homology identified with HMMER3 (REF. 115) to predict the impact of a missense variant
PROVEAN ¹¹⁶	Missense and indel prediction	Y (N)	Y	N	The degree of protein sequence conservation is used to predict the impact of an amino acid change or an indel
REVEL ¹¹⁷	Missense prediction (ensemble method)	Y (Y)	N	N	Incorporates 18 individual scores from 13 different tools to produce an ensemble 'pathogenicity' score for missense variants
PhastCons ¹¹⁸	Sequence conservation	Y (N)	Y	Y	Uses multiple sequence alignments from diverse species to identify conserved elements
PhyloP ^{92,118}	Sequence conservation	Y (N)	Y	Y	Uses multiple sequence alignments from diverse species to assign per-base P values of conservation
GERP++ ⁹³	Sequence conservation	Y (N)	Y	Y	Measures sequence conservation in the human genome through alignments to 43 other vertebrate genomes
MutationTaster2 (REF. 119)	Multi-data integration	Y (N)	Y	Y (intronic)	Integrates sequence conservation, as well as data from the 1000 Genomes Project, ENCODE ⁹⁰ and ClinVar, to predict the consequence of variants within a gene model
VAAST ^{6,24,66}	Multi-data integration	Y (N)	Y	Y	Integrates variant frequency data with phylogenetic conservation for variant prioritization and burden testing
CADD ³⁹	Multi-data integration	Y (N)	Y (short indels)	Y	Integration of conservation metrics, functional data (for example, DNase I hypersensitivity and transcription factor binding) and scores such as SIFT and PolyPhen2 to predict the deleteriousness of nucleotide or short indel change in the genome
FitCons ⁴⁰	Multi-data integration	Y (N)	Y (short indels)	Y	Integrates functional genomic data from the ENCODE project to cluster genomic regions and to predict the probability of a fitness consequence based on sequence conservation and the degree of regional polymorphism in the human genome

Tableau 9. Les outils bioinformatiques de prédiction de l'impact fonctionnel des variations faux-sens les plus fréquemment utilisés (d'après Eilbeck *et al.*, 2017)

Pour une même variation, les prédictions obtenues peuvent différer selon l'outil utilisé (Thursberg *et al.*, 2011). Il est donc essentiel de combiner et d'interroger plusieurs outils de prédiction afin de bénéficier des avantages de chaque programme en minimisant les limites de ceux-ci de façon à obtenir la prédiction la plus fiable possible. C'est en ce sens que sont désormais développés des outils dits « intégrés », tel que l'algorithme REVEL (Ioannidis *et al.*, 2016), visant à regrouper les prédictions émises par divers algorithmes afin d'augmenter la fiabilité des prédictions *in silico*. La spécificité de ces outils n'est cependant pas encore suffisante (Miosge *et al.*, 2015 ; Kerr *et al.*, 2017) et, aujourd'hui, les prédictions bioinformatiques de l'impact fonctionnel des variations ne peuvent aucunement être utilisées seules pour statuer sur le caractère neutre ou pathogène des variations (Richards *et al.*, 2015). Toutefois, elles peuvent constituer un argument de classification agrégé à d'autres éléments de preuve. De façon importante, ces approches *in silico* peuvent être mises à profit dans le but de présélectionner les variations les plus susceptibles d'avoir un impact fonctionnel afin de les inclure dans des essais fonctionnels expérimentaux.

Acquisition et utilisation des données fonctionnelles

Les études expérimentales fonctionnelles représentent un argument puissant dans la démonstration de l'impact pathogène ou neutre d'une variation (Tableau 7 ; Richards *et al.*, 2015). Là encore, il est important de distinguer au moins deux types d'effet soit au niveau de l'ARN soit au niveau de la protéine, voire la combinaison des deux.

Les anomalies d'épissage de l'ARN induites par certaines variations sont principalement recherchées à l'aide d'analyses expérimentales conduites à partir de l'ARN du patient ou sur la base de l'utilisation de minigènes (approches détaillées dans la partie IV.3.). Ces approches sont universelles car déclinables aisément pour l'étude du profil d'épissage de la plupart des gènes.

A l'inverse, les investigations menées au niveau protéique sont souvent spécifiques d'une fonction donnée et les approches méthodologiques sont donc très diverses. De fait, elles permettent d'interroger directement la fonction ou le mécanisme susceptible d'être altéré par la variation d'intérêt, dans un contexte *in vivo* ou *in vitro*. Dans le cas de protéines remplissant différentes fonctions ou impliquées dans plusieurs voies biologiques, il s'agit d'évaluer l'activité la plus en lien avec l'étiologie de la maladie considérée. Plus le test analyse la fonction biologique complète de la protéine au sein d'un environnement reflétant celui d'origine, plus le résultat obtenu pourra être considéré comme un argument fort pour l'interprétation de la variation. Il est à noter que la robustesse des données, la reproductibilité du test, mais également la calibration de l'essai constituent des conditions indispensables à l'utilisation de ces résultats comme arguments de classification. Ces critères peuvent donc représenter une limite à l'utilisation de certains tests sans constituer toutefois les seules contraintes associées à certaines de ces analyses. En effet, certaines de ces approches fonctionnelles sont développées à partir du matériel biologique du patient et celui-ci n'est pas toujours disponible. De plus, il existe rarement un test fonctionnel standard unique capable de tester toutes les fonctions biologiques d'une protéine, rendant ainsi nécessaire d'effectuer plusieurs tests fonctionnels dans le but d'obtenir un résultat conclusif. Ces analyses particulièrement chronophages ne peuvent donc être réalisées systématiquement et dans un contexte de routine au sein d'un laboratoire de diagnostic. Les données fonctionnelles ne sont donc que rarement disponibles malgré leur haut potentiel informatif.

La collecte de l'ensemble de ces données constitue, par un effet combinatoire et additif, un faisceau d'arguments permettant l'interprétation précise d'une variation génétique. Toutefois, ces arguments sont tous de natures différentes et sont ainsi associés à des avantages et des inconvénients qui leur sont propres. L'ACMG-AMP a donc proposé d'accorder à chaque argument un poids, plus ou

moins important, en fonction de sa nature. Deux classes de critères ont été établies : (i) ceux en faveur du caractère neutre de la variation tels que BA (*stand-alone*) désigne des arguments individuels suffisants, BS (*Benign Strong*) des arguments forts et BP (*Benign Supporting*) des arguments faibles ; et (ii) ceux en faveur de la pathogénicité de la variation tels que PVS (*Pathogenic Very Strong*) désigne des arguments très forts, PS (*Pathogenic Strong*) des arguments forts, PM (*Pathogenic Moderate*) des arguments modérés et PP (*Pathogenic supporting*) des arguments faibles (Tableau 7 ; Richards *et al.*, 2015). Chaque argument est ainsi pondéré et combiné au sein d'un algorithme décisionnel reposant sur un système bayésien, permettant d'établir une probabilité de certitude que la variation soit pathogène ou bénigne. Ainsi, une probabilité de neutralité supérieure à 99% sera associée à une variation de classe 1 et entre 90% et 99%, de classe 2, tandis qu'une probabilité de pathogénicité entre 90% et 99% sera associée à une variation de classe 4 et supérieure à 99%, de classe 5. Cette classification reflète alors le poids combiné de l'ensemble des arguments finalement collectés.

Néanmoins, ces recommandations restent très générales et, en l'absence de données suffisantes, elles ne permettent pas toujours d'affirmer ou d'infirmer la pathogénicité d'une variation, laissant ainsi une part non négligeable de variations de signification inconnue (VSI, classe 3) inutilisables pour le conseil génétique, et ce malgré les efforts permanents mis en œuvre dans le but de faire évoluer ces recommandations (Kim *et al.*, 2019). Afin d'accompagner les laboratoires de diagnostic dans leur démarche, de nombreuses sociétés savantes à l'échelle nationale et internationale se sont constituées en vue d'établir des critères toujours plus précis d'interprétation des variations, notamment en émettant des recommandations spécifiques dans le contexte d'une maladie donnée.

II.2. Les recommandations spécifiques dans le contexte du syndrome sein-ovaire

L'identification de la variation pathogène est primordiale dans un contexte évocateur du syndrome sein-ovaire pour l'optimisation de la prise en charge médicale des patients et des apparentés asymptomatiques, non seulement d'un point de vue diagnostique et pronostique mais également thérapeutique, comme explicité précédemment (voir partie I.1.4.). La problématique de l'interprétation des variations dans le contexte du syndrome sein-ovaire est de ce fait très prégnante d'autant plus que les gènes majeurs de prédisposition, *BRCA1* et *BRCA2*, présentent un très large spectre mutationnel (Caputo *et al.*, 2012 ; Cline *et al.*, 2018).

Depuis l'introduction en 1996 du séquençage de ces 2 gènes dans la routine des laboratoires de diagnostic, des millions de patients et leurs apparentés ont bénéficié de ce criblage moléculaire dans le monde. Ces analyses ont permis d'identifier plusieurs milliers de variations différentes dans

ces gènes. Face à cette pléthore de variations et à la nécessité de délivrer une interprétation rigoureuse de chacune d'entre elles dans un contexte de génétique médicale, les laboratoires se sont regroupés en réseaux nationaux et internationaux afin de partager ces informations et leurs expertises spécifiques dans le contexte du syndrome sein-ovaire. En France, la constitution d'un groupe de travail dans ce domaine a été initiée au sein du groupe Génétique et Cancer (GGC, <http://www.unicancer.fr/en/cancer-and-genetic-group>). Ce réseau national collige l'ensemble des acteurs de santé nationaux issus de diverses structures (CHU, hôpitaux publics et privés) et disciplines (cliniciens, biologistes, conseillers, chercheurs) dans le domaine de l'oncogénétique afin d'évaluer les risques familiaux de cancer, d'élaborer les modalités de prise en charge des familles et de contribuer à l'amélioration des connaissances sur les prédispositions génétiques aux cancers. En 2003, le GGC a fait l'objet d'un des trois plans cancer élaboré par l'INCa permettant ainsi la constitution d'un dispositif national d'oncogénétique regroupant aujourd'hui 26 laboratoires en charge de la réalisation des tests génétiques et 148 sites de consultations et ce, afin d'offrir une prise en charge équitable à l'ensemble de la population. Les missions de ce groupe sont multiples et variées telles que la coordination des projets d'évaluation des bonnes pratiques cliniques et de recherche, la constitution de cohortes épidémiologiques, la réalisation d'essais cliniques ou encore, l'établissement d'études visant à identifier de nouveaux gènes de prédisposition au syndrome sein-ovaire et l'expertise des variations génétiques identifiées en France. On peut notamment citer un projet majeur de ce groupe, l'étude COVAR (COségréagation VARIants), qui a pour objectif de collecter les données de co-ségrégation familiale associées aux VSI dans le but de contribuer à la classification des variants *BRCA1/BRCA2*.

Une des initiatives remarquables du groupe GGC a été l'établissement d'une base de données spécifique au syndrome sein-ovaire : BRCA Share™ *BRCA1/BRCA2*, initialement nommée UMD-*BRCA1/BRCA2* (<http://www.umd.be/BRCA1> ; <http://www.umd.be/BRCA2>) (Caputo *et al.*, 2012 ; Bérout *et al.*, 2016). L'objectif de cette base de données est non seulement d'agrèger l'ensemble des variations identifiées dans le génome des patients au sein des 16 laboratoires français en charge du diagnostic moléculaire de ce syndrome, mais également de collecter les informations génétiques, cliniques et biologiques associées telles que les données de co-occurrence, de co-ségrégation, tumorales et fonctionnelles (Caputo *et al.*, 2012). Sous sa forme actuelle accessible en ligne, cette base répertorie environ 2 500 et 3 500 variations différentes dans les gènes *BRCA1* et *BRCA2*, respectivement (Tableau 10). La plupart de ces changements de séquence correspondent à des variations ponctuelles. Le caractère pathogène a pu être établi pour environ 30% de l'ensemble des variations répertoriées (36% pour *BRCA1* et 28% pour *BRCA2*) sur la base des données collectées et du système de classification ACMG. Mais, pour plus de 50% d'entre elles (54% pour *BRCA1* et 63% pour *BRCA2*), à défaut de données suffisamment informatives, il n'a pas été possible de statuer sur le

caractère pathogène ou neutre et, dans ce cas, la classification VSI a été retenue pour ces variations qui, pour la plupart, sont très rares (Tableau 10).

BRCA1	All phenotypes		
	All records	Different variants	Samples
TOTAL	11824	2495	10691
Large rearrangements	441	77	440
Large deletions (>=1 exon)	350	59	350
Large duplications (>=1 exon)	91	18	90
Small rearrangements	9247	1812	8603
Small deletions (<1 exon)	2134	372	2128
Small insertions (<1 exon)	915	132	914
Splice sites (<10 bp from exon)	597	133	592
Point mutations	5366	1136	4931
Nonsense	1202	191	1197
Missense	4164	945	3795
Indels	235	39	235
Mid-intronic variations	2136	606	1946

Biological significance			
	All records	Different variants	Samples
1 - Neutral	3038	133	2632
2 - Likely Neutral	702	113	685
3 - VUS	2414	1356	2294
4 - Likely causal	13	2	13
5 - Causal	5659	891	5637

BRCA2	All phenotypes		
	All records	Different variants	Samples
TOTAL	15955	3454	13450
Large rearrangements	39	26	39
Large deletions (>=1 exon)	37	25	37
Large duplications (>=1 exon)	2	1	2
Small rearrangements	12573	2739	11052
Small deletions (<1 exon)	2311	480	2303
Small insertions (<1 exon)	563	160	560
Splice sites (<10 bp from exon)	724	158	713
Point mutations	8885	1907	7773
Nonsense	1051	220	1048
Missense	7834	1687	6854
Indels	90	34	90
Mid-intronic variations	3343	689	3075

Biological significance			
	All records	Different variants	Samples
1 - Neutral	4816	134	4053
2 - Likely Neutral	2356	171	2205
3 - VUS	4577	2181	4278
4 - Likely causal	0	0	0
5 - Causal	4223	968	4185



Tableau 10. Types et nombre de variations répertoriées dans les bases de données BRCA Share™ BRCA1/BRCA2 (d'après <http://www.umd.be/BRCA1> ; <http://www.umd.be/BRCA2>). Décembre 2020

Ces bases de données traduisent ainsi l'ampleur du défi de l'interprétation clinique des variations dans le contexte du syndrome sein-ovaire et l'obstacle majeur qu'il constitue à l'optimisation de la prise en charge médicale des patients et de leurs apparentés. Cette problématique est particulièrement importante pour certains types de variants. Ainsi, il est à noter que près de 75% des variations faux-sens identifiées dans le gène *BRCA2* sont classées VSI (Figure 19 ; Starita *et al.*, 2017).

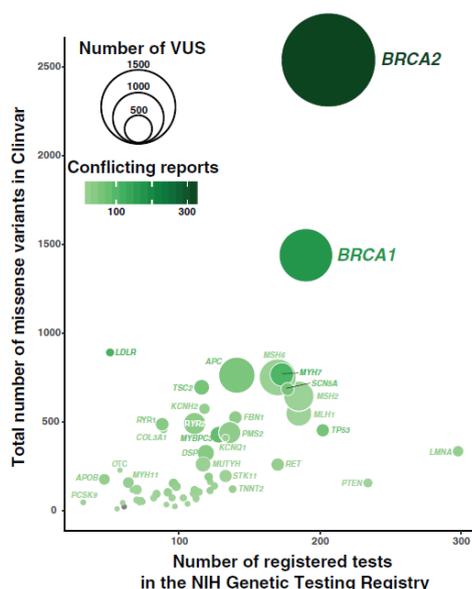


Figure 19. La majorité des variations faux-sens identifiées dans le gène *BRCA2* sont des VSI (d'après Starita *et al.*, 2017). Le nombre de tests enregistrés corrèle avec le nombre de variations faux-sens identifiées. Les données intégrées dans ce graphique ont été extraites de la base de données ClinVar le 5 avril 2017.

Au niveau international, plusieurs groupes de recherche multidisciplinaires se forment afin de classer les variations identifiées dans les gènes de prédisposition au syndrome sein-ovaire. ENIGMA (*Evidence-based Network for the Interpretation of Germline Mutant Alleles*, <https://enigmaconsortium.org/>) est un consortium d'experts rassemblant ces différents groupes de travail dans le but de fournir un effort de classification collaboratif à l'échelle internationale. En se basant sur le travail de l'IARC publié en 2008 par Plon et ses collaborateurs (Plon *et al.*, 2008), le consortium ENIGMA a également établi des recommandations standardisées de classification spécifiques des variations identifiées dans les gènes *BRCA* selon cinq classes de pathogénicité (Spurdle *et al.*, 2012). Les critères de classification pris en compte correspondent à ceux établis sur la base des recommandations générales de l'ACMG-AMP (Richards *et al.*, 2015), mais ils intègrent également des données issues de stratégies de classification spécifiquement développées dans le contexte du syndrome sein-ovaire. Ainsi, plusieurs modèles multifactoriels de rapport de vraisemblance ont été développés pour appréhender la pathogénicité des variations de *BRCA1* et *BRCA2*. Ces approches multifactorielles intègrent dans leur calcul les informations relatives à la nature de la variation, sa localisation, sa fréquence dans la population générale, l'histoire personnelle et familiale de cancers des patients porteurs, les données tumorales, la co-ségrégation, la co-occurrence de variants pathogènes en *trans* dans le même gène (a priori exclue puisque associée à un phénotype d'anémie de Fanconi), ainsi que les prédictions bioinformatiques et les analyses fonctionnelles (Goldgar *et al.*, 2004 ; Easton *et al.*, 2007, Tavtigian *et al.*, 2008 ; Parsons *et al.*, 2019 ; Li *et al.*, 2020). Tout comme pour les recommandations établies par l'ACMG-AMP, chaque argument est associé à une pondération. Cependant, le poids de chaque argument est ici spécifique de l'interprétation des variations dans les gènes *BRCA*, permettant ainsi une analyse plus fine des variations identifiées dans un contexte de syndrome sein-ovaire. Par exemple, toute variation associée à une fréquence supérieure à 1% dans la population générale sera classée comme variation neutre (classe 1). Dans certains cas, il est nécessaire de prendre en compte l'ethnicité de la patiente avant d'interpréter cette donnée. En effet, certaines populations présentent une prévalence élevée de certaines variations pathogènes dans les gènes *BRCA* du fait d'effet fondateur, par exemple dans la population ashkénaze ou polonaise (Roa *et al.*, 1996 ; Gorski *et al.*, 2004). L'ensemble des recommandations standardisées de classification des variations identifiées dans les gènes *BRCA* établies par ENIGMA est listé dans le tableau 11. De façon remarquable, le consortium ENIGMA ainsi que d'autres équipes internationales ont contribué à l'évaluation d'outils de prédiction et d'essais fonctionnels spécifiques de l'impact des variations *BRCA1* et *BRCA2* sur le processus d'épissage mais également sur différentes fonctions de ces protéines (Kuznetsov *et al.*, 2008, 2010 ; Hendriks *et al.*, 2014 ; Mesman *et al.*, 2019, 2020 ; Canson *et al.*, 2020 ; Tubeuf *et al.*, 2020a, 2020b). Ces études ont permis d'établir les performances de chaque outil ou test expérimental et d'intégrer ces informations dans la méthode de classification de ces variations. La

validation et la calibration de ces approches *in silico* et fonctionnelles constituent un apport considérable pour la classification des VSI de *BRCA1/2*.

Tableau 11. Critères de classification des variations identifiées dans les gènes *BRCA* selon le consortium ENIGMA (d'après les guidelines v.2.5.1 2017 du consortium Enigma, https://enigmaconsortium.org/wp-content/uploads/2020/08/ENIGMA_Rules_2017-06-29-v2_5_1.pdf)

Class	Criterion
Class 5: pathogenic	Posterior probability of pathogenicity >0.99 from multifactorial likelihood analysis.
	Coding sequence variant encoding a premature termination codon i.e. nonsense/frameshift predicted to disrupt expression of clinically important functional domain(s)/residue(s).
	The variant allele produces only transcripts that lead to a premature stop codon, or in-frame deletion predicted to disrupt clinically important domains, as determined by RNA assays on patient germline tissue that assess allele-specific transcript expression.
	Copy number deletion removing exon(s) spanning clinically important residue(s) or proven to result in a frameshift alteration predicted to interrupt expression of clinically important residue(s).
	Copy number duplication proven to result in frameshift alteration predicted to interrupt expression of clinically important residue(s).
Class 4: likely pathogenic	In the absence of clinical evidence to assign an alternative classification, sequence variant that alters the <i>BRCA1</i> or <i>BRCA2</i> translation initiation methionine site.
	Variant at IVS±1 or IVS±2 or G>non-G at last base of exon when adjacent intronic sequence is not GTRRGT that is predicted to alter used of native donor/acceptor site AND is untested for splicing aberrations using RNA assays on patient blood that assess allele-specific transcript expression, AND is not predicted or known to alter production of (naturally occurring) in-frame RNA isoforms that may rescue gene functionality.
	A variant that encodes the same amino acid change as a previously established Class 5 pathogenic missense variant with a different underlying nucleotide change, is located in a known clinically important functional protein domain, with no evidence of mRNA aberration (splicing or expression) from in vitro mRNA assays on patient RNA, and the variant is absent from outbred control reference groups.
	A small in-frame deletion variant that removes a codon for which a missense substitution Class 5 pathogenic variant has been described, is located in a known clinically important functional protein domain, unlikely to result in an alternative aberration via mRNA splicing, and is absent from outbred control reference

Class 3: uncertain	Posterior probability of pathogenicity 0.05-0.949 from multifactorial likelihood analysis.
	In the absence of clinical evidence to assign an alternative classification, variant allele tested for mRNA aberrations using <i>in vitro</i> assays of patient RNA that assess allele-specific transcript expression, and is found to produce mRNA transcript(s) predicted to encode intact full-length protein and/or isoforms that do not disrupt known clinically important functional residue(s).
	Insufficient evidence to classify variant.
	Variant located at position listed in Table 6, unless proven to fall in another class based on additional evidence.
	Variant with conflicting evidence for pathogenicity.
Class 2: Likely not pathogenic or of little clinical significance	Posterior probability of pathogenicity 0.001-0.049 from multifactorial likelihood analysis.
	An exonic variant, that encodes the same amino acid change as a previously established Class 1 not pathogenic missense variant with a different underlying nucleotide change, and for which there is low bioinformatic likelihood to disrupt normal splicing.
	Synonymous substitution with low bioinformatic likelihood to disrupt normal splicing, determined to have combined prior probability of pathogenicity ≤ 0.02 from clinically calibrated bioinformatic analyses.
Class 1: not pathogenic or of no clinical significance	Posterior probability of pathogenicity < 0.001 from multifactorial likelihood analysis
	Variant with reported frequency $\geq 1\%$ in large outbred control reference groups
	Exonic variant that encodes the same amino acid change as a previously established Class 1 Not Pathogenic missense variant with a different underlying nucleotide change, and for which there is no evidence of mRNA aberration from <i>in vitro</i> mRNA assays.
	Exonic variant encoding a missense substitution or a small in-frame insertion/deletion with prior probability ≤ 0.02 from clinically calibrated bioinformatic analyses OR intronic variant AND Low bioinformatic likelihood to disrupt normal splicing OR Increased bioinformatic likelihood to disrupt normal splicing but no associated mRNA aberration (splicing or allelic imbalance) as determined using <i>in vitro</i> laboratory assays AND
	Co-occurrence <i>in trans</i> with a known pathogenic sequence variant in the same gene in an individual with no obvious additional clinical phenotype other than BRCA-associated cancer OR Allele frequency ≥ 0.001 and < 0.01 in large outbred control reference groups.

Fort de son expertise internationale, le consortium ENIGMA a contribué à la création de la base de données *BRCA* Exchange (<https://brcaexchange.org>) (Cline *et al.*, 2018). Aujourd'hui, c'est plus de 40 000 variations ponctuelles identifiées dans les gènes *BRCA* qui ont été partagées dans *BRCA* Exchange à partir de l'agrégation des informations reportées dans de multiples bases de données en lien avec le phénotype associé au syndrome sein-ovaire parmi lesquels ENIGMA mais aussi LOVD, ClinVar et BIC (*Breast cancer Information Core database* ; <https://research.nhgri.nih.gov/bic/>) (Szabo *et al.*, 2000). La base *BRCA* Exchange intègre également les données issues de bases populationnelles telles que celles générées dans le cadre du projet 1000 Génomes ou celles compilées par gnomAD (Cline *et al.*, 2018). *BRCA* Exchange permet ainsi de mutualiser les données très exhaustives de toutes les variations identifiées dans les gènes *BRCA* à un niveau international. Cette base propose de surcroît, pour près de 7 250 mutations, une interprétation établie par des experts du domaine (Cline *et al.*, 2018).

II.3. Les recommandations spécifiques dans le contexte du syndrome de Lynch

Au même titre que le syndrome sein-ovaire, l'identification d'une variation pathogène dans l'un des gènes *MMR* dans un contexte évocateur du syndrome de Lynch oriente la prise en charge médicale du patient et de sa famille, avec des implications importantes au niveau des options thérapeutiques, des possibilités de chimioprévention et de chirurgies prophylactiques ainsi que de la surveillance des apparentés asymptomatiques (Vasen *et al.*, 2013) (voir chapitre I.2.3.). L'interprétation de ces variations représente donc un enjeu majeur en génétique médicale mais également un défi étant donné le très large spectre mutationnel des gènes impliqués. Depuis 1995, plusieurs dizaines de milliers de cas index évocateurs d'un syndrome de Lynch ont été adressés à l'un des 16 laboratoires du réseau français dans le but de bénéficier d'un criblage moléculaire des gènes *MMR* (Grandval *et al.*, 2013). Ces analyses ont permis d'identifier plus d'un millier de variations différentes, principalement dans les gènes *MLH1* et *MSH2* (Tableau 12). Les gènes *MMR* étant des gènes suppresseurs de tumeurs au même titre que les gènes *BRCA*, les altérations constitutionnelles hétérozygotes capables de conduire à leur déficience sont de mêmes natures que celles identifiées dans le cas d'une déficience des gènes *BRCA*. Cela peut s'agir de larges réarrangements de types insertions, duplications, délétions, mais dans la majorité des cas il s'agit de variations ponctuelles telles que de petites insertions/délétions, des variations non-sens, faux-sens, synonymes ou encore, des variations introniques sur les sites d'épissage (Plazzer *et al.*, 2013).

MLH1	All phenotypes		
	All records	Different variants	Samples
TOTAL	3063	568	2527
Large rearrangements	133	41	132
Large deletions (>=1 exon)	100	29	99
Large duplications (>=1 exon)	33	12	33
Small rearrangements	2332	455	2182
Small deletions (<1 exon)	453	92	450
Small insertions (<1 exon)	51	26	51
Splice sites (<10 bp from exon)	354	83	345
Point mutations	1412	245	1310
Nonsense	293	47	290
Missense & Synonymous	1119	198	1029
Mid-intronic variations	598	72	523
Indels	62	9	60

Biological significance of non-truncating variations			
	All records	Different variants	Samples
1 - Neutral	545	41	525
2 - Likely Neutral	23	15	23
3 - UV	1166	257	1001
4 - Likely causal	72	23	68
5 - Causal	1257	232	1239

MSH2	All phenotypes		
	All records	Different variants	Samples
TOTAL	3231	638	2646
Large rearrangements	391	57	389
Large deletions (>=1 exon)	350	41	348
Large duplications (>=1 exon)	41	16	41
Small rearrangements	2479	529	2158
Small deletions (<1 exon)	322	109	319
Small insertions (<1 exon)	109	28	109
Splice sites (<10 bp from exon)	867	77	669
Point mutations	1156	307	1120
Nonsense	379	77	379
Missense & Synonymous	777	230	745
Mid-intronic variations	361	53	324
Indels	25	8	24

Biological significance of non-truncating variations			
	All records	Different variants	Samples
1 - Neutral	212	29	205
2 - Likely Neutral	23	14	23
3 - UV	1486	288	1004
4 - Likely causal	44	16	44
5 - Causal	1466	291	1461



Tableau 12. Types et nombre de variations répertoriées dans les bases de données UMD-*MLH1/MSH2* (d'après <https://www.umd.be/MLH1> ; <https://www.umd.be/MSH2>). Décembre 2020

Ces variations sont répertoriées dans les bases de données nationales UMD-*MLH1/MSH2/MSH6* (<https://www.umd.be/MLH1> ; <https://www.umd.be/MSH2> ; <https://www.umd.be/MSH6>) développées par le GGC dans le but de compiler l'ensemble des variations génétiques identifiées dans les trois gènes majeurs de prédisposition au syndrome de Lynch dans le génome de patients suivis en France (Grandval *et al.*, 2013). A ce jour, ce sont près de 1 650 variations qui ont été rapportées dans ces trois bases de données, conjointement à leurs informations génétiques (données de co-occurrence, de ségrégation, cliniques, tumorales et fonctionnelles). La collecte de ces informations par le GGC et l'expertise de ce groupe de travail dans le domaine de l'interprétation des variations localisées dans les gènes *MMR* a permis d'établir le caractère pathogène pour 37% à 46% des variations identifiées. Toutefois, près de la moitié des variations rapportées dans ces bases (45% à 49%, selon le gène) restent encore classées VSI à défaut d'informations suffisantes (Tableau 12).

Au niveau international, l'organisation multidisciplinaire InSiGHT (*International Society for Gastrointestinal Hereditary Tumours*), fondée en 2005, est constituée d'experts ayant pour objectif d'agrèger les connaissances de l'ensemble des prédispositions aux cancers héréditaires gastro-intestinaux, dont notamment le syndrome de Lynch (Plazzer *et al.*, 2013). Fort d'une expertise dans le domaine de l'interprétation des variations acquise au fil des années, InSiGHT a formé en 2011 un comité de 45 experts (VIC, *Variant Interpretation Committee*) dans le but d'établir ces propres recommandations de classifications des variations identifiées dans les gènes *MMR* et ce, afin de

standardiser l'interprétation des variations dans ces gènes (Tableau 13 ; Thompson *et al.*, 2014). Au même titre que les recommandations émises par le consortium ENIGMA, celles de l'organisation InSiGHT sont basées sur les critères précédemment établis par l'IARC et reposent sur une analyse multifactorielle incluant la nature de la variation, sa localisation, sa fréquence dans la population générale, le contexte familial, les données tumorales, de ségrégation, de co-occurrence ainsi que les analyses bioinformatiques et fonctionnelles (Tableau 13). Les arguments pris en compte sont de même nature que ceux considérés par l'ACMG dans le contexte d'une maladie à déterminisme mendélien (Richards *et al.*, 2015) mais des éléments spécifiques au syndrome de Lynch sont également agrégés pour l'interprétation des variations dans les gènes *MMR*. Il s'agit notamment (i) de l'instabilité microsatellitaire tumorale (statut MSI), signature d'une défaillance au niveau du système MMR, (ii) de la perte d'expression des protéines MMR recherchée par immunohistochimie dans les tumeurs et (iii) de la co-occurrence en *trans* de variations pathogènes dans les gènes *MMR* *a priori* exclue puisqu'associée au syndrome CMMRD (Thompson *et al.*, 2014).

Tableau 13. Critères de classification des variations identifiées dans les gènes *MMR* selon l'organisation InSiGHT (d'après les guidelines v.2.4. 2018 de l'organisation InSiGHT, https://www.insight-group.org/content/uploads/2018/08/2018-06_InSiGHT_VIC_v2.4.pdf)

Class 5 – Pathogenic
<ul style="list-style-type: none"> • Variants with probability of pathogenicity >0.99 (based on co-segregation in more than one family, and without conflicting tumour data) using a multifactorial likelihood model • Variants in the initiation codon of MLH1 • Coding sequence variation resulting in a stop codon i.e. a nonsense or frameshift alteration that is not after codon 743 in MLH1 or after codon 888 in MSH2, and not in the last exon of MSH6 or PMS2*. For variants after codon 743 in MLH1 or after codon 888 in MSH2, or in the last exon of MSH6 or PMS2, use other criteria to classify. • Variants where mRNA assays using RNA derived from patient constitutional biological samples indicate that the variant allele results in a splicing aberration (with evidence that the variant allele produces no full-length/reference transcript) leading to premature stop codon or in-frame deletion disrupting a functional domain or protein conformation. Splicing aberration must be confirmed in a minigene assay or an additional RNA assay from an independent laboratory if it is not a predicted splice site mutation. • Large genomic deletions • Large genomic duplications shown by laboratory studies (which define the breakpoints of the duplication) to result in a frameshift before the last splice junction • Variants demonstrating all of the following characteristics with no conflicting results (combined evidence achieves estimated LR >100:1, posterior >0.99 with prior 0.5): ✓ Variant-specific abrogated function in protein or mRNA based lab assays (see MMR functional assay supplementary material, Figure 1) <ul style="list-style-type: none"> - OR - co-occurrence <i>in trans</i> with a known pathogenic sequence variant in the same gene in a patient with clinical features consistent with Constitutional Mismatch Repair Deficiency (CMMRD) and documented MMR deficiency in normal cells ^b - OR - presence of the variant on different haplotypes across families indicating that reported LS clinical features are not due to an undiscovered sequence change <i>in cis</i> with the variant ✓ Evidence for co-segregation with disease where pedigree information provided allows calculation of likelihood ratio (LR) of $\geq 10:1$ ^c or >5 informative meioses <ul style="list-style-type: none"> - OR - at least one revised Amsterdam criteria¹⁹ family with ≥ 4 affected carriers - OR - across ≥ 2 revised Amsterdam criteria families reported to show segregation with disease (no further information provided) - OR - ≥ 2 families with ≥ 3 affected non-proband carriers in each family ✓ ≥ 2 independent MSI-H tumours using a standard panel of 5-10 markers ^d and/or loss of MMR protein expression consistent with the variant location (may include tumour information from proband) ✓ evidence that variant is not an undescribed polymorphism at allele frequency >1% (minor allele frequency [MAF] >0.01) in appropriate population control reference groups ^e

Class 4 – Likely pathogenic

- Variants with probability of pathogenicity between 0.95-0.99 (or >0.99 if co-segregation in one family and presence of conflicting tumour data) using a multifactorial likelihood model
- Variants in the initiation codon of MSH6 or PMS2
- Variants with monoallelic expression - complete loss of expression of the variant allele (full-length transcript should be analysed with and without NMD block)
- De novo with maternity and/or paternity unconfirmed in a case with MMR deficient (MSI/immunoloss consistent with affected gene, with no MLH1 methylation) Lynch spectrum tumour^f
- Variants at IVS±1 or IVS±2^g
- Variants demonstrating (combined evidence achieves LR >20:1, posterior 0.95-0.99 with prior 0.5):
 - ✓ variant-specific abrogated function in protein or mRNA based lab assays (see MMR functional assay supplementary material)
 - OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with clinical features consistent with CMMRD and documented MMR deficiency in normal cells^b
 - OR - presence of the variant on different haplotypes across families indicating that reported LS clinical features are not due to an undiscovered sequence change *in cis* with the variant
 - ✓ plus one of the following:
 - co-segregation with disease assessed by available pedigree information allows calculation of LR of >5:1^g or >3 informative meioses
 - OR - at least one revised Amsterdam criteria family with ≥3 affected carriers
 - OR - ≥2 families with ≥2 affected non-proband carriers
 - OR - ≥2 independent tumours with MSI using a standard panel of 5-10 markers^d and/or loss of MMR protein expression consistent with the variant location (may include tumour information from proband)
- A missense substitution that encodes the same amino acid change with a different underlying nucleotide change as a previously established Class 5 pathogenic missense variant with normal RNA result^h, and is absent from appropriate population control reference groups^e
 - * Otherwise, if the previously established Class 5 pathogenic missense variant truly is a splice defect, the new missense variant also has to be investigated on a functional level for RNA splicing.

Class 3 – Uncertain

- Variants with probability of pathogenicity between 0.05-0.949 using a multifactorial likelihood model
- Variants in the initiation codon of MSH2
- Variants that have insufficient evidence (molecular or otherwise) to classify, which may include large genomic duplications not yet shown by laboratory studies to result in a frameshift before the last splice junction, missense alterations, small in-frame insertions/deletions, silent variants, intronic variants, promoter and regulatory region variants

Conflicting lines of evidence (e.g. IHC/MSI does not fit, segregation does not fit, etc)

Class 2 – Likely not pathogenic/little clinical significance

- Variants with probability of pathogenicity between 0.001-0.049 using a multifactorial likelihood model
- Synonymous substitutions and intronic variants with no associated mRNA aberration (either splicing or allelic imbalance) as determined by laboratory assays conducted with nonsense-mediated decay inhibition. Whenever abnormal transcripts are identified at similar levels in controls they will be considered naturally occurring isoforms and not mRNA aberrations
- Variants reported to occur in a specific ethnic group at allele frequency ≥1%, (MAF ≥ 0.01, tested in ≥ 160 individuals) and that have not yet been excluded as known founder pathogenic sequence variants ("founder mutations")
- 5' UTR or promoter region variants which have bi-allelic expression, shown by cDNA experiment that both alleles are expressed ~ 50/50 including one with the variant or at least one heterozygous exonic variant in the transcript (full-length transcript should be analysed with and without NMD block)
- Variants demonstrating:
 - ✓ variant-specific proficient function in protein and mRNA based lab assays (see MMR functional assay supplementary material)
 - OR - co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in a patient with colorectal cancer after age 45 (or other LS cancer above the median age of onset for that cancer in LSⁱ), and who has no previous or current evidence of clinical manifestations of CMMRD
 - ✓ plus one of the following:
 - present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)
 - OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01^j)
 - OR - estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures
 - OR - ≥3 colorectal tumours with MSS and/or no loss of MMR protein expression and/or loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation
- Variants demonstrating any three of the following:
 - present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals)
 - OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR≤ 0.01^j)
 - OR - estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures
 - OR - ≥3 colorectal tumours with MSS and/or no loss of MMR protein expression and/or loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation
- An exonic variant, that encodes the same amino acid change as a previously established Class 1 missense variant with a different underlying nucleotide change, and for which there is no evidence of mRNA aberration (splicing or allelic imbalance) as determined using *in vitro* laboratory assays conducted with nonsense-mediated decay inhibition

<p>Class 1 – Not pathogenic/no clinical significance</p> <ul style="list-style-type: none"> • Variants with probability of pathogenicity <0.001 using a multifactorial likelihood model • Variants reported to occur in control reference groups at allele frequency $\geq 1\%$ (MAF ≥ 0.01, tested in ≥ 160 individuals) and excluded as founder pathogenic sequence variants • 5' UTR or promoter region variants demonstrating: <ul style="list-style-type: none"> ✓ Bi-allelic expression, shown by cDNA experiment that both alleles are expressed – 50/50 including one with the variant or at least one heterozygous exonic variant in the transcript (full-length transcript should be analysed with and without NMD block) ✓ Meets Class 1 Frequency criteria: <ul style="list-style-type: none"> - present in control reference groups at allele frequency $\geq 0.01\%$ (MAF ≥ 0.0001, tested in ≥ 160 individuals) <ul style="list-style-type: none"> - OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR ≤ 0.01 °) • Variants demonstrating: <ul style="list-style-type: none"> ✓ Variant-specific proficient function in protein and mRNA based lab assays (see MMR functional assay supplementary material) - OR - co-occurrence <i>in trans</i> with a known pathogenic sequence variant in the same gene in a patient with colorectal cancer after age 45 (or other LS cancer above the median age of onset for that cancer in LS °), and who has no previous or current evidence of clinical manifestations of CMMRD ✓ Plus any two of the following: <ul style="list-style-type: none"> - present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals) - OR - lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR ≤ 0.01 °) - OR - estimated risk with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures - OR - ≥ 3 colorectal tumours with MSS and/or no loss of MMR protein expression and/or loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation • Variants demonstrating all of the following: <ul style="list-style-type: none"> ✓ Present in control reference groups at allele frequency 0.01-1% (MAF 0.0001-0.01, tested in ≥ 160 individuals) ✓ Lack of co-segregation with disease consistent with a dominant high-risk pathogenic sequence variant in pedigrees (LR ≤ 0.01 °) ✓ Estimated risk <1.5, with upper bound 95% confidence limit <4, as determined by large well-designed case-control studies that consider size, geography/ethnicity and quality control measures ✓ ≥ 3 colorectal tumours with MSS and/or no loss of MMR protein expression and/or loss of MMR protein(s) in LS spectrum tumours, that is inconsistent with the gene demonstrating genetic variation

Il est à noter que les recommandations d'InSiGHT ont été établies avant celles d'ENIGMA et ont donc constituées une source d'inspiration importante pour la création de ces dernières. Afin de favoriser le partage des connaissances et notamment les informations ne faisant pas toujours l'objet de publications mais pouvant servir d'arguments dans la classification des variations, l'organisation InSiGHT a développé une large base de données publique regroupant plus de 2 300 variations identifiées dans les gènes associés aux formes de prédisposition héréditaire de cancers gastro-intestinaux (<https://insight-database.org>) (Plazzer *et al.*, 2013 ; Thompson *et al.*, 2014). Ainsi, cette base de données permet d'agrèger l'ensemble des informations associées à une variation identifiée au sein des gènes *MMR*. Même si la stratégie spécifique d'interprétation des variations localisées dans un des gènes *MMR* a permis de proposer une classification rigoureuse pour un grand nombre de variations, il n'en reste pas moins qu'environ 30% de ces modifications génétiques sont encore classées VSI (Thompson *et al.*, 2014).

Malgré l'ensemble de ces recommandations ainsi que tous les efforts nationaux et internationaux fournis, le statut pathogène ou bénin de très nombreuses variations identifiées dans les gènes *BRCA* et *MMR* reste encore indéterminé et constitue un obstacle majeur à la prise en charge des familles porteuses de ces VSI. L'interprétation biologique d'une variation constitue un élément important en vue de sa classification clinique et relève bien souvent de l'évaluation de sa capacité présumée à altérer la structure ou la fonction de la protéine via des analyses expérimentales et

bioinformatiques. Toutefois, il est actuellement recommandé qu'une analyse des transcrits doit précéder les analyses protéiques pour toutes variations capables d'altérer l'épissage telles que les variations introniques, faux-sens, synonymes et les insertions et délétions en phase, selon le comité d'interprétation des variations d'InSiGHT et ENIGMA. En effet, il est maintenant bien établi qu'une fraction importante des variations pathogènes identifiées dans les gènes associés à des maladies monogéniques sont des variations capables d'altérer un processus en amont de la traduction, l'épissage des ARN pré-messagers (Baralle *et al.*, 2009).

III. L'épissage, étape clé de la maturation des ARN pré-messagers

Il y a maintenant près de 60 ans que la découverte de l'acide ribonucléique messager (ARNm) a révolutionné notre compréhension de l'expression génique (Cobb, 2015). Cette découverte a permis d'établir les bases de la théorie fondamentale de la biologie moléculaire en positionnant l'ARNm comme l'intermédiaire essentiel entre l'ADN, support de l'information génétique, et la protéine, effectrice d'une fonction cellulaire spécifique. Il faudra ensuite attendre plus de dix ans pour comprendre que l'ARNm n'est pas une simple copie par transcription du gène mais le résultat d'un processus complexe de maturation d'un ARNm précurseur (pré-ARNm). Trois étapes sont alors à distinguer avant l'export de l'ARNm mature dans le cytoplasme : l'ajout d'une coiffe en 5', la polyadénylation en 3' et l'épissage.

L'épissage est un mécanisme co-transcriptionnel clé de la maturation du pré-ARNm consistant, d'une part, en l'élimination des régions introniques et, d'autre part, en l'assemblage des séquences exoniques consécutives (Cartegni *et al.*, 2002). La découverte de l'existence d'un tel processus fut inhérente à la mise en évidence, très inattendue à l'époque, de la différence structurelle entre l'ARN nucléaire et l'ARN cytoplasmique mais également, de la discontinuité des gènes. C'est ainsi, qu'en 1977, les équipes de Richard Roberts, biochimiste britannique, et de Phillip Sharp, biochimiste américain, démontrèrent expérimentalement pour la première fois la discontinuité des gènes dans un modèle d'adénovirus (Berget *et al.*, 1977 ; Chow *et al.*, 1977 ; Berk, 2016). L'observation par la technique de microscopie électronique de molécules hybrides ADN/ARN cytoplasmique a permis de montrer que si certaines régions étaient complémentaires, d'autres avaient complètement disparu de l'ARNm cytoplasmique, supportant ainsi l'hypothèse qu'une partie du matériel génétique est excisée de l'ARNm avant l'export de celui-ci dans le cytoplasme (Figure 20). Ces travaux ont contribué de façon déterminante à la définition des gènes intégrant les notions d'intron/exon et d'épissage. Ces concepts relatifs à l'expression génique se sont révélés universels chez les eucaryotes et cette découverte fondamentale a été récompensée par un prix Nobel de Médecine en 1993 (Carr, 1993). La recherche, ces 40 dernières années, a permis (i) la dissection des mécanismes biochimiques sous-tendant la réaction catalytique d'épissage, (ii) l'identification des différents acteurs moléculaires composant le spliceosome, (iii) le décryptage, au moins en partie, des différents signaux d'épissage qui constituent le code d'épissage et (iv) la caractérisation de l'épissage alternatif comme régulateur de l'expression génique.

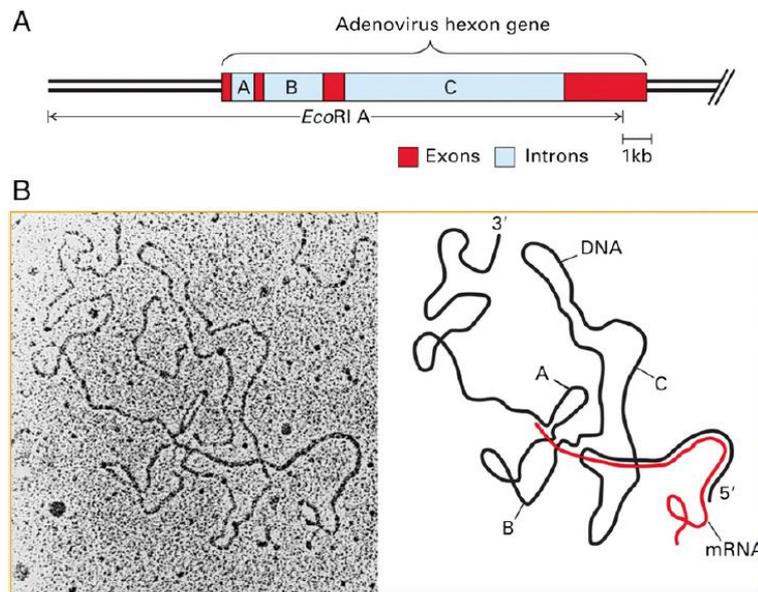


Figure 20. Mise en évidence de molécules hybrides ADN/ARN par la technique de microscopie électronique (d'après Berk, 2016 adapté de Berget *et al.*, 1977). A. Représentation schématique du gène hexon de l'adénovirus. B. Dans le panel de droite, image de microscopie électronique et dans le panel de gauche, représentation schématique de l'hybridation entre les molécules d'ARNm en rouge et d'ADN en noir.

III.1. La réaction catalytique d'épissage

La réaction d'épissage s'effectue en deux étapes de transestérification conduisant, d'une part, à l'excision de l'intron sous la forme d'une structure en lasso et, d'autre part, à la ligation des exons entre eux (Figure 21 ; Cech, 1986 ; Patel and Steitz, 2003 ; Shi, 2017a). La première réaction de transestérification a lieu entre le groupement 2'-OH de l'adénine du site de branchement (BS, *Branch Site*), située au niveau intronique, et le phosphate de la guanine du site 5' d'épissage en amont du BS. Cette attaque nucléophile permet de couper l'extrémité 5' de l'intron tout en établissant une liaison phosphodiester 2'-5' intra-intronique (structure en lasso) ce qui libère l'extrémité 3'-OH de l'exon en amont. La deuxième réaction de transestérification consiste à l'attaque nucléophile de la liaison phosphodiester au niveau du site 3' d'épissage (en aval) par ce groupement 3'OH. Cette réaction permet de couper l'extrémité 3' de l'intron et de lier ensemble les deux exons consécutifs de façon à produire un ARNm mature. L'intron est libéré sous la forme d'un lasso puis clivé par l'enzyme de débranchement DBR1 et finalement dégradé par des ribonucléases.

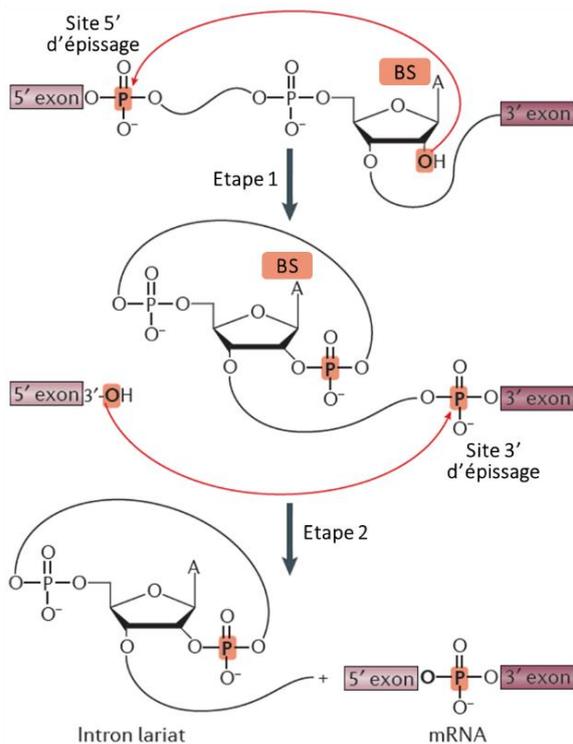


Figure 21. Les réactions de transestérification du pré-ARNm (adapté de Shi, 2017a). Les étapes 1 et 2 correspondent aux première et deuxième réactions de transestérifications, respectivement.

D'un point de vue biochimique, la réaction d'épissage semble simple mais l'exécution de celle-ci repose sur une machinerie enzymatique complexe et dynamique : le splicéosome, une macromolécule de plusieurs mégadaltons.

III.2. La machinerie d'épissage : le splicéosome

Le splicéosome est un complexe ribonucléoprotéique dont le réarrangement constant de sa composition et de son réseau d'interactions, au cours de sa progression dans le cycle de réactions catalytiques, confère précision et flexibilité à la machinerie d'épissage. Ce complexe est composé de cinq petites ribonucléoprotéines nucléaires (snRNP, *small nuclear RiboNucleoProteins*) et d'un très large nombre de facteurs protéiques non-snRNPs (150-300) (Reed, 1996 ; Will and Lührmann, 2011 ; Shi, 2017a ; Dufner-Almeida *et al.*, 2019). Chaque snRNP est composée d'une petite molécule d'ARN non codant (snRNA, *small nuclear RNA*) d'environ 90 à 220 nucléotides, associée à une dizaine de protéines différentes dont certaines sont communes à toutes les snRNPs, comme les protéines Sm, et d'autres sont spécifiques à chaque snRNP. C'est au cours des années 1980 que le développement d'essais *in vitro* d'épissage basés sur l'utilisation d'extraits nucléaires actifs s'est révélé essentiel pour la caractérisation fonctionnelle des différents acteurs du splicéosome (Padgett *et al.*, 1983 ; Hernandez and Keller, 1983). Ainsi, en formant des interactions ARN-ARN par appariements de bases, les snRNPs sont capables d'interagir directement avec le pré-ARNm au niveau de séquences spécifiques correspondant à des signaux d'épissage, mais également, entre elles (Nilsen, 1994). De plus, les snRNPs

sont responsables du recrutement de facteurs protéiques spécifiques, via des interactions ARN-protéines, lesquelles, avec l'ensemble des protéines non-snRNPs, facilitent les interactions protéine-protéine, ARN-protéine et ARN-ARN au niveau du spliceosome (Konaska and Sharp, 1987). Au cours des dix dernières années, la purification des différents composants du spliceosome ainsi que leur analyse par microscopie électronique ont permis de remarquables progrès quant à la caractérisation des propriétés biochimiques associées à chaque complexe du spliceosome, résultat des interactions séquentielles des snRNP et des facteurs protéiques avec le pré-ARNm et entre eux (Will and Lührmann, 2011 ; Wahl *et al.*, 2009). Il est également apparu que le centre catalytique du spliceosome est directement composé d'ARN, traduisant ainsi la nature ribozyme du spliceosome (Fica *et al.*, 2013 ; Shi *et al.*, 2017b).

Dans les cellules eucaryotes, deux types de machinerie d'épissage co-existent : (i) le spliceosome majeur, responsable de l'épissage de la très grande majorité des introns, les introns dits majeurs ou de type U2 et (ii) le spliceosome mineur, à l'origine de l'épissage d'introns atypiques, les introns dits mineurs ou de type U12.

III.2.1. Le spliceosome majeur

Le spliceosome majeur, responsable de l'épissage de 99% des pré-ARNm, est composé des 5 snRNPs U1, U2, U4, U5 et U6. Il est également assisté de plusieurs facteurs protéiques non-snRNPs dont notamment des protéines de liaison à l'ARN (RBP, *RNA Binding Protein*) telles que U2AF et SF1, des protéines riches en arginine et sérine (SR, *Serine aRginine-rich*) ainsi que diverses enzymes dont notamment, des hélicases, des kinases et des phosphatases (Shi, 2017a). Dans les cellules eucaryotes, le spliceosome majeur est capable de reconnaître les signaux d'épissage au niveau du pré-ARNm différemment selon la taille des introns, définissant ainsi deux modèles de reconnaissance : le modèle de définition d'intron et celui de définition d'exon (Figure 22 ; Berget *et al.*, 1995 ; Fox-Walsh *et al.*, 2005). Dans le cas d'introns de petite taille, entre 200 et 250 nucléotides, l'appariement des composants du spliceosome aura lieu au niveau des signaux d'épissage de part et d'autre de l'intron (modèle de définition d'intron) tandis que dans le cas d'introns de grande taille, le spliceosome reconnaîtra spécifiquement les signaux d'épissage flanquant un exon (modèle de définition d'exon). Les signaux d'épissage principaux (sites 5' et 3' d'épissage, site de branchement, région riche en pyrimidine) et auxiliaires (éléments *cis* régulateurs de l'épissage) sont détaillés plus en amont dans la partie III.3.

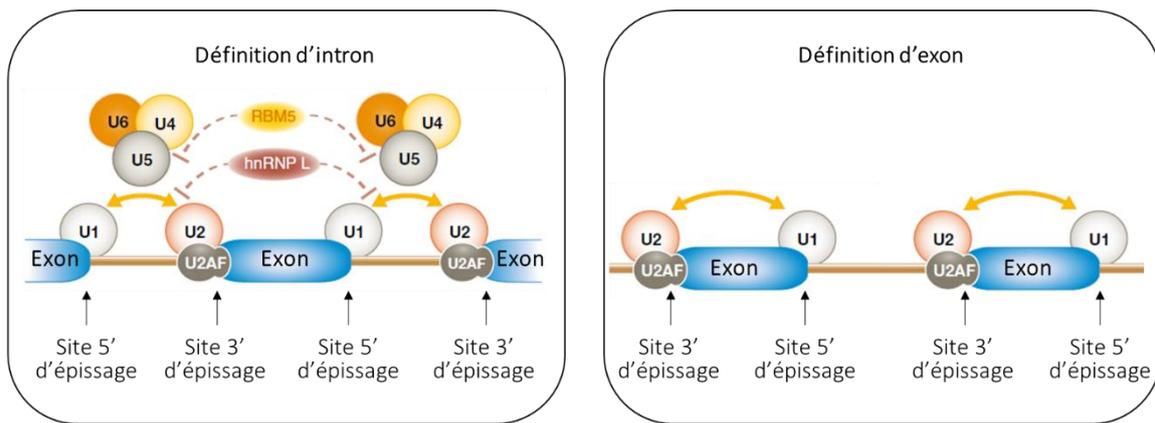


Figure 22. Modèles de reconnaissance des signaux d'épissage par le splicéosome majeur (adapté de Dagenet *et al.*, 2015).

Dans le modèle de définition d'intron, la première étape d'assemblage du splicéosome majeur au niveau du pré-ARNm conduit à la formation du complexe E, résultat de l'interaction de la snRNP U1 avec le site 5' d'épissage ainsi que de la fixation de facteurs protéiques non-snRNPs, la molécule SF1 (*Splicing Factor 1*) aussi appelée mBPP (*mammalian Branch Point binding Protein*) et l'hétérodimère U2AF65/U2AF35 (U2AF, *U2 Auxillary Factor*), au niveau du site de branchement et de la région riche en pyrimidine, respectivement (Figure 23 ; Matera and Wang, 2014 ; Dufner-Almeida *et al.*, 2019). Dans un second temps, la formation du complexe A, également appelé pré-splicéosome, consiste en le recrutement de la snRNP U2 au niveau du site de branchement, nécessitant alors la libération de la molécule SF1 et la stabilisation de l'interaction à l'aide des facteurs protéiques SF3a et SF3b. Par la suite, les snRNPs U4, U5 et U6 s'assemblent pour former une tri-snRNP avant de rejoindre le complexe A, générant ainsi le complexe B qui est pré-catalytique, c'est-à-dire qu'il n'est pas encore actif. Son activation est accomplie à la suite de réarrangements dans les interactions snARN-ARN et ARN-protéines qui à leur tour engendrent une déstabilisation des interactions avec les snRNP U1 et U4 conduisant ainsi à leur libération. Le splicéosome devient alors actif par l'intervention de l'hélicase Brr2 (complexe B^{act}), puis par l'hélicase Prp2 (complexe B*), ce qui conduit à la première réaction de transestérification. Une fois celle-ci achevée, le complexe B* devient le complexe C qui, après plusieurs réarrangements et l'action de l'hélicase Prp16, est catalytiquement actif sous la forme du complexe C*, à l'origine de la seconde réaction de transestérification. Pour finir, le splicéosome se dissocie libérant ainsi l'ARNm mature, à l'aide de l'hélicase Prp22, ainsi que l'ensemble des composants du splicéosome qui participeront à d'autres réactions d'épissage.

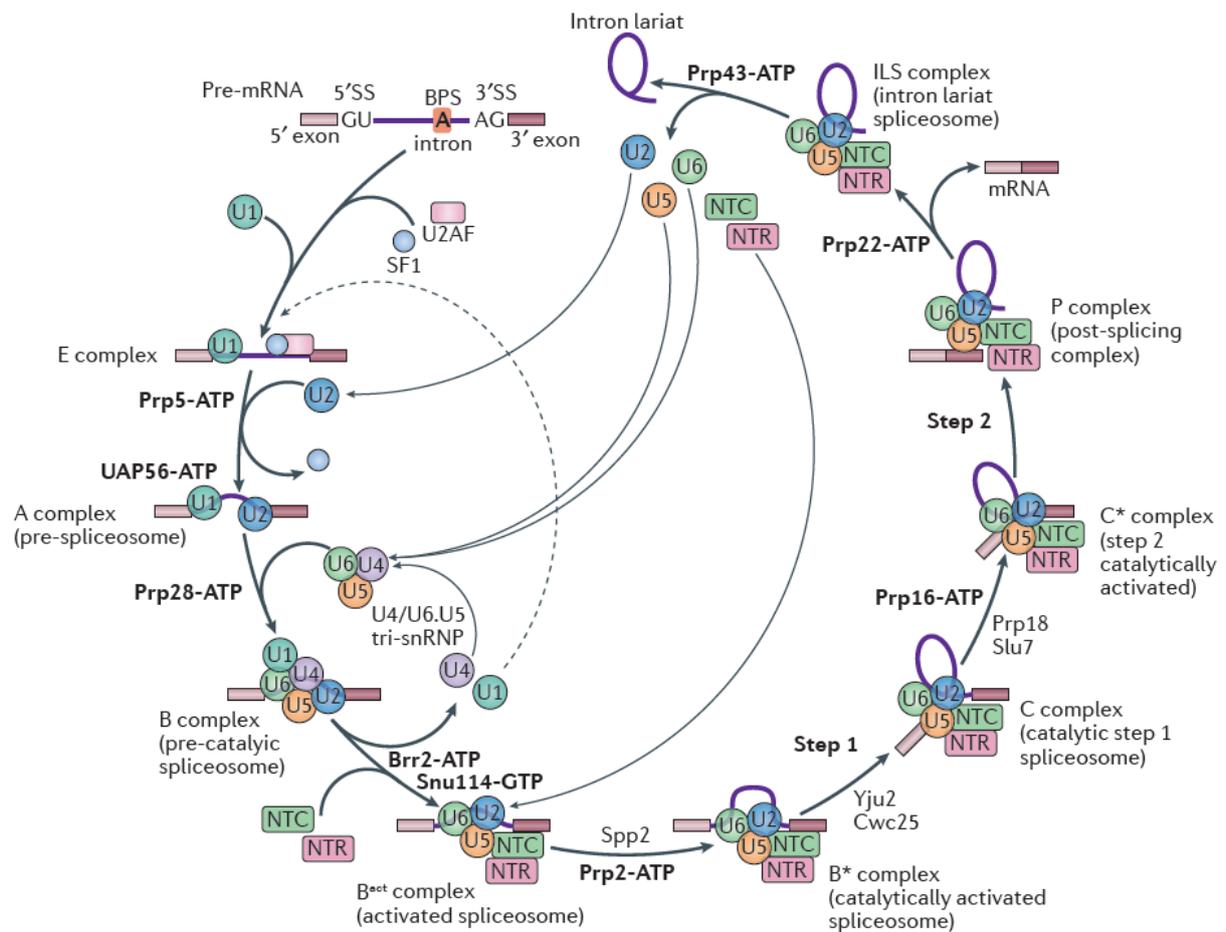


Figure 23. Les étapes d'assemblage du splicéosome majeur selon le modèle de définition d'intron (d'après Shi, 2017a)

Le modèle de définition d'intron a largement été établi en utilisant comme modèle d'étude la levure car son génome est majoritairement constitué de petits introns séparant de larges exons. A l'inverse, dans la grande majorité des gènes humains, les introns sont de grande taille et séparent de petits exons. Ainsi, le modèle de définition d'exon est celui le plus couramment utilisé chez l'Homme. Dans ce modèle, les snRNP U2AF et U1 se lient à la région riche en pyrimidine et au site 5' d'épissage, respectivement, qui caractérisent un même exon (Figure 22 ; Daguenet *et al.*, 2015). Le recrutement de la snRNP U2 au niveau du site de branchement est ici simultané à celui des protéines SR au niveau des éléments exoniques activateurs d'épissage. Ensemble, ces protéines forment un réseau de stabilisation dans l'exon. D'autres facteurs protéiques peuvent également être amenés à être recrutés selon la nature des éléments de régulation présents dans l'exon et les introns adjacents. Par la suite, le modèle de définition d'exon est converti en modèle de définition d'intron et se déroule comme décrit ci-dessus (Will and Lührmann, 2011 ; Matera and Wang, 2014). Les mécanismes à l'origine de cette transition restent encore mal définis. Certaines hypothèses sous-tendent que le modèle de définition d'exon pourrait se convertir en complexe A ou encore pourrait recruter la tri-snRNP U4/U6.U5 et former directement le complexe B (Li *et al.*, 2019).

III.2.2. Le splicéosome mineur

Le second type de machinerie d'épissage, le splicéosome mineur, est à l'origine de l'épissage de moins de 1% des pré-ARNm. Ce second splicéosome est constitué des 5 snRNPs U11, U12, U4atac, U6atac et U5, seule snRNP commune aux deux complexes (Schneider *et al.*, 2002 ; Patel and Steitz, 2003 ; Will and Lührmann, 2005). Parmi les quelques centaines de protéines non-snRNPs associées à la machinerie d'épissage, seules sept d'entre elles sont spécifiquement associées au splicéosome mineur. La composition hautement similaire des deux machineries d'épissage se reflète également dans l'assemblage de celles-ci, malgré quelques différences, notamment dans l'étape de reconnaissance des signaux d'épissage. En effet, dans les introns mineurs, c'est le dimère de snRNPs U11/U12 (analogue du dimère U1/U2) qui est à l'origine de la reconnaissance coopérative du site 5' d'épissage et du site de branchement du pré-ARNm, formant ainsi directement le complexe A sans passer par le complexe E. Dans un second temps, le complexe B résulte du recrutement de la tri-snRNP U4atac/U6atac.U5 et, au même titre que le splicéosome majeur, les snRNPs U11 et U4atac sont déstabilisées à la suite de réarrangements et libérées afin d'activer le complexe B*, responsable de la catalyse de la première réaction de transestérification. Enfin, le complexe C-like généré après cette première réaction assure la seconde réaction de transestérification avant dissociation complète du splicéosome et libération de la molécule d'ARNm mature selon la même cascade d'évènements que pour le splicéosome majeur.

III.3. Les signaux d'épissage

Le splicéosome, qu'il soit majeur ou mineur, reconnaît précisément les bornes délimitant les introns et les exons sur la base de la présence de plusieurs motifs au niveau des pré-ARNms. Il s'agit de séquences spécifiques correspondant notamment aux signaux d'épissage principaux (sites 5' et 3' d'épissage, site de branchement, région riche en pyrimidine) et auxiliaires (éléments *cis* régulateurs de l'épissage).

III.3.1. Les signaux principaux

Ce n'est qu'un an seulement après la découverte de l'épissage que l'équipe de Pierre Chambon, biochimiste et généticien français, rapporte dans le cadre de l'étude du gène de l'ovalbumine, l'existence de séquences très similaires localisées systématiquement au niveau des jonctions exon/intron et potentiellement associées à un rôle majeur dans la reconnaissance du pré-ARNm par la machinerie d'épissage (Breathnach *et al.*, 1978). Dès lors, plusieurs études ont démontré, par alignement de séquences, que les jonctions exon/intron sont reconnues par le splicéosome sur la

base de séquences spécifiques consensus, les sites 5' et 3' d'épissage (5'/3'ss, *splice site*), également appelés sites donneurs et accepteurs d'épissage, respectivement (Figure 24 ; Mount, 1982 ; Senapathy *et al.*, 1990 ; Cartegni *et al.*, 2002 ; Will and Lührmann, 2011). La séquence des 5'ss inclut les trois dernières bases de l'exon et les six premières bases de l'intron tandis que la séquence des 3'ss comprend les vingt dernières bases de l'intron et les trois premières bases de l'exon (Yeo and Burge, 2004 ; Frederiks *et al.*, 2015). Selon la nature de l'intron, U2 ou U12, les séquences consensus de ces sites d'épissage ne présentent pas la même combinaison de nucléotides.

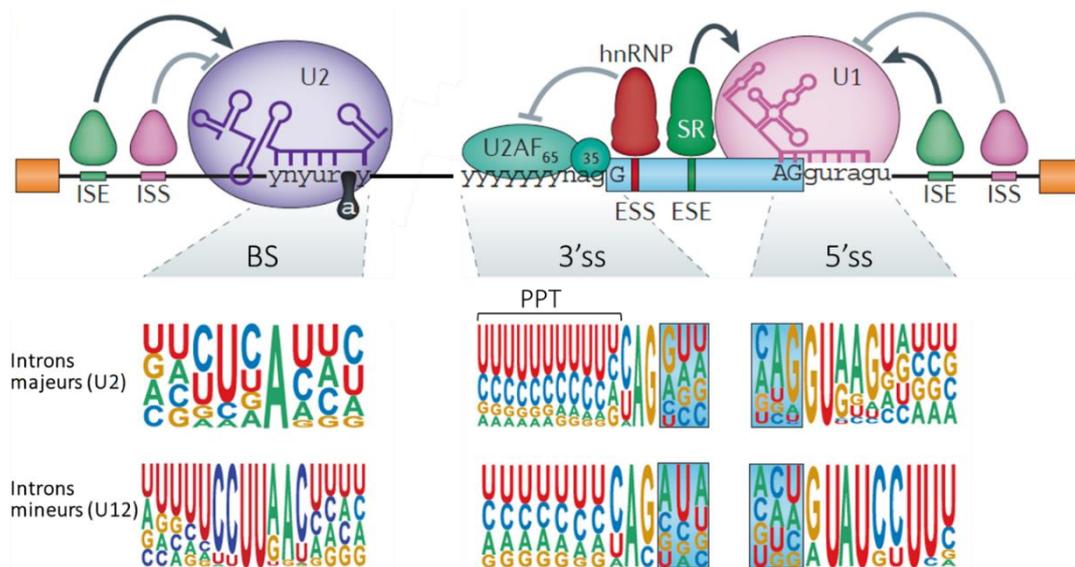


Figure 24. Représentation schématique des signaux d'épissage principaux (adapté de Scotti and Swansson, 2016). ISE, *Intronic Splicing Enhancer* ; ISS, *Intronic Splicing Silencer* ; ESE, *Exonic Splicing Enhancer* ; ESS, *Exonic Splicing Silencer* ; BS, *Branch Site* ; PPT, *PolyPyrimidine Tract* ; 3'ss, *3' splice site* ; 5'ss, *5' splice site*.

Dans les introns majeurs ou introns U2, le motif consensus des 5'ss est CAG|GURAGU, où « | » représente la jonction exon/intron, et le motif plus conservé des 3'ss est YAG|G où « | » indique la jonction intron/exon. Les dinucléotides introniques en gras, GU en position +1 et +2 pour le 5'ss et AG en position -2 et -1 pour le 3'ss, correspondent aux dinucléotides les plus invariables des sites d'épissage (Figure 24 ; Cartegni *et al.*, 2002). En effet, ces dinucléotides introniques caractérisent près de 99% des introns U2, alors couramment appelés introns GU-AG. Les autres introns U2 sont, quant à eux, définis par des sites d'épissage dits atypiques dont les plus fréquents sont les introns GC-AG qui représentent près de 0,9% des sites d'épissage (Burset *et al.*, 2000 ; Thanaraj and Clark, 2001 ; Sheth *et al.*, 2006 ; Sibley *et al.*, 2016). Dans ces introns, il apparait que la substitution d'un T vers un C en position +2 au niveau du 5'ss conduit à un mésappariement avec la snRNP U1 qui, même s'il est compensé par le reste du motif, tend à affaiblir les sites donneurs d'épissage des introns GC-AG

comparativement à ceux des introns GU-AG. Il est d'ailleurs intéressant de noter que les 5'ss présentant les dinucléotides GC sont plus favorablement associés à des introns alternatifs (Thanaraj and Clark, 2001).

La reconnaissance des 3'ss dans les introns majeurs met également en jeu d'autres éléments : la région riche en pyrimidine et le site de branchement (Figure 24 ; Gao *et al.*, 2008 ; Mercer *et al.*, 2015). La région riche en pyrimidine (PPT, *PolyPyrimidine Tract*) comme son nom l'indique est une séquence caractérisée par une haute densité de pyrimidines. Elle se situe dans la séquence consensus du 3'ss, entre 10 à 12 nucléotides en amont de la jonction intron/exon. Le site de branchement (BS, *Branch Site*) quant à lui, est une séquence consensus courte localisée, dans la majorité des BS humains, entre 18 et 44 nucléotides en amont du site 3'ss. Des travaux se sont penchés sur la cartographie expérimentale du BS d'un nombre limité d'introns mais ce n'est que très récemment qu'une étude a décrit expérimentalement, pour la première fois à large échelle, les BS d'un très grand nombre d'introns grâce à une technique de séquençage à haut débit de l'ARN lasso (Mercer *et al.*, 2015). Les auteurs de cette étude ont ainsi pu déterminer le motif consensus du point de branchement CUGAC, où l'adénosine est le nucléotide le plus hautement conservé car elle correspond au point de branchement (BP, *Branch Point*) indispensable à la première réaction de transestérification.

Les introns mineurs, ou introns U12, présentent des caractéristiques distinctes des introns majeurs (Figure 24 ; Patel and Steitz, 2003 ; Will and Lührmann, 2005). A l'origine, ces introns mineurs ont été identifiés sur la base de la différence majeure de combinaison des dinucléotides introniques des sites d'épissage : AU en position +1 et +2 pour le 5'ss et AC en position -2 et -1 pour le 3'ss. Cependant, il est apparu par la suite que les introns mineurs pouvaient également être définis par les dinucléotides GU-AG, voir même d'autres combinaisons de nucléotides telles que AU-AA et AU-AG. Il apparaît donc que la différence de composition des dinucléotides introniques localisés dans les sites d'épissage ne représente pas un élément déterminant dans la distinction des introns mineurs. En effet, et contrairement aux introns majeurs, les introns dits mineurs se définissent par une région moins riche en pyrimidine associée au 3'ss. En outre, le point de branchement des introns mineurs est particulièrement conservé et localisé entre 11 à 13 nucléotides en amont du 3'ss, consistant ainsi un facteur important pour leur identification par la machinerie d'épissage. Même si ces introns sont très peu fréquents, leur épissage est indispensable à la survie cellulaire car ils sont notamment présents dans des gènes assurant des fonctions essentielles telles que la réplication et la réparation de l'ADN, la transcription ou encore la traduction (Patel and Steitz, 2003).

III.3.2. Les signaux auxiliaires

L'information contenue dans les principaux signaux est indispensable pour la reconnaissance des jonctions exon/intron par le splicéosome mais reste insuffisante pour la définition précise de certains exons. En effet, le recrutement et l'activité du splicéosome peuvent être modulés par de nombreux éléments *cis* régulateurs de l'épissage (SRE, *Splicing Regulatory Elements*) qui, contrairement aux signaux d'épissage précédemment cités, restent très peu caractérisés. Ils sont généralement décrits comme de courts motifs de 6 à 8 nucléotides permettant l'interaction des pré-ARNm avec des facteurs de régulation spécifiques favorisant ou inhibant l'inclusion de l'exon en agissant notamment sur l'activité du splicéosome (Cartegni *et al.*, 2002 ; Wang and Burge, 2008). Les SRE peuvent se distinguer selon leur localisation sur le pré-ARNm mais également selon leur rôle. Ainsi, certains éléments peuvent être introniques (ISR, *Intronic Splicing Regulatory elements*) tandis que d'autres sont exoniques (ESR, *Exonic Splicing Regulatory elements*) (Figure 24). Les éléments introniques de régulation de l'épissage restent aujourd'hui encore moins bien caractérisés que les éléments exoniques. La plupart des études concentrent leurs analyses sur les éléments de régulation localisés entre 200 à 300 nucléotides autour des sites d'épissage, même si l'existence d'éléments de régulation à plus grande distance a pu être démontrée (Fu and Ares, 2014). Chacun de ces éléments est potentiellement associé à une activité stimulatrice (*enhancer*) ou inhibitrice (*silencer*) de l'épissage. Différentes approches expérimentales et bioinformatiques ont été mises à profit pour permettre l'identification et la caractérisation fonctionnelle des SRE (Liu *et al.*, 1998 ; Fairbrother *et al.*, 2002 ; Smith *et al.*, 2006 ; Goren *et al.*, 2006 ; Ke *et al.*, 2011). Ainsi, l'équipe du Dr Chasin a développé une méthode appelée QUEPASA (*Quantifying extensive phenotypic arrays from sequence arrays*) qui a permis l'évaluation exhaustive et quantitative de l'activité ESR de tous les hexamères possibles (4 096 motifs) (Ke *et al.*, 2011). L'effet sur l'épissage de chaque hexamère a été quantifié dans un essai basé sur l'utilisation de minigènes, sous la forme d'un score (ESRseq) rendant compte de l'activité potentielle *enhancer* (ESE) ou *silencer* (ESS) de chaque motif. Il est toutefois important de noter que le rôle des ESR est dépendant de leurs localisations par rapport aux sites d'épissage et de la séquence environnante. En effet, ces éléments parfois chevauchants fonctionnent selon un mode combinatoire coopératif ou compétitif (Ke *et al.*, 2011).

Les éléments exoniques activateurs d'épissage (ESE, *Exonic Splicing Enhancer*) sont présents dans la majorité des exons à des densités variables (Cartegni *et al.*, 2002 ; Wang and Burge, 2008). Ils sont essentiellement reconnus par des facteurs de régulation de l'épissage appartenant à la famille des protéines SR qui se fixent spécifiquement aux motifs ESE présents sur le pré-ARNm (Figure 24), via une interaction avec leur extrémité N-terminale (Graveley, 2000 ; Shepard and Hertel, 2009). Ces

protéines ubiquitaires sont capables d'interagir avec de nombreux partenaires protéiques, via leur domaine riche en arginine et sérine localisé au niveau de l'extrémité C-terminale. Elles peuvent notamment se lier aux protéines constituant le splicéosome ce qui leur permet de réguler positivement l'inclusion de l'exon. En outre, le domaine riche en arginine et sérine permet également de réguler l'activité et la localisation des protéines SR via leur phosphorylation. Dans le noyau, les protéines SR sont transportées vers les sites actifs de transcription où elles vont faciliter l'assemblage du complexe E du splicéosome (Graveley *et al.*, 2000).

Les éléments exoniques inhibiteurs de l'épissage (ESS, *Exonic Splicing Silencer*), quant à eux, permettent la fixation de facteurs de régulation de l'épissage favorisant l'exclusion de l'exon tels que les protéines hnRNP (*heterogeneous nuclear RiboNucléoProteins*) (Figure 24 ; Cartegni *et al.*, 2002 ; Wang and Burge, 2008). Ces protéines se fixent aux ESS grâce à leur motif de reconnaissance de l'ARN et peuvent également contenir des domaines d'inhibition de l'épissage tels que des motifs riches en glycine (Pozzoli and Sironi, 2005 ; Martinez-Contreras, 2007). Comme les protéines SR, l'activité des protéines hnRNP est médiée par leur interaction avec d'autres facteurs protéiques dont notamment d'autres protéines hnRNP. Ainsi, en formant un complexe répressif, les protéines hnRNP induisent une inhibition de l'épissage soit en empêchant la reconnaissance des sites d'épissage soit en agissant comme antagonistes des facteurs activateurs d'épissage ou des composants du splicéosome.

L'épissage met donc en jeu des éléments *cis* (signaux d'épissage principaux et auxiliaires) et des facteurs *trans* (splicéosome et facteurs de régulation) qui permettent une régulation complexe et très fine de ce processus. La complexité de celle-ci est illustrée notamment par la spécificité tissulaire de certains épissages alternatifs (Dvinge, 2018) mais également par les modifications de l'épissage observées lors du processus de tumorigenèse (Escobar-Hoyos *et al.*, 2019).

III.4. L'épissage alternatif de l'ARNm

Longtemps, il a été considéré qu'un gène permettait la production d'une protéine unique et était ainsi associé à une fonction cellulaire spécifique. Or, le génome humain est constitué d'approximativement 20 000 à 25 000 gènes codant pour des protéines et conduit pourtant à la production d'un nombre de protéines distinctes très largement supérieur (> 100 000) (International Human Genome Sequencing Consortium, 2004). Ce paradoxe s'explique, en partie, du fait de l'existence d'un mécanisme rendant possible la production de plusieurs protéines à partir d'un même gène : l'épissage alternatif. Chez les eucaryotes, ce processus permet l'expansion du répertoire protéique ou protéome en conduisant potentiellement à la production, à partir d'un unique transcrit

primaire (pré-ARNm), d'une multitude d'ARNm mature aboutissant éventuellement à tout autant de protéines (Matlin *et al.*, 2005 ; Nilsen and Graveley, 2010). En effet, si certains exons sont systématiquement retrouvés dans la molécule d'ARNm considérée comme l'isoforme de référence et résultant d'un épissage dit constitutif, de nombreux exons sont, quant à eux, alternativement épissés. Un exemple saisissant est celui du gène *Dscam* chez l'espèce *Drosophila melanogaster* qui conduit à la production de 38 016 ARNm différents à partir de la même population de pré-ARNm (Schmucker *et al.*, 2000). Chez l'Homme, il est estimé que plus de 95% des gènes multi-exons humains conduisent à la production de transcrits alternatifs (Pan *et al.*, 2008). En moyenne, un gène humain est associé à 2,53 isoformes d'épissage qui sont potentiellement traduites en protéines (Tress *et al.*, 2007).

La multiplicité des ARNm produits par épissage alternatif peut être due à des évènements multiples et non-exclusifs, tels que : (i) le saut d'exons cassettes, qui se caractérise par la perte d'inclusion d'un ou plusieurs exons dans le transcrit mature ; (ii) l'utilisation d'un site 3' d'épissage alternatif qui conduit soit à une rétention intronique en amont de l'exon, soit à une excision de la partie proximale de l'exon ; (iii) l'utilisation d'un site 5' d'épissage alternatif qui conduit soit à une rétention intronique en aval de l'exon, soit à une excision de la partie terminale de l'exon ; (iv) des exons alternatifs mutuellement exclusifs (si l'un est incorporé dans l'ARNm mature, l'autre ne sera pas présent et inversement) ; et (v) la rétention complète d'un intron (Figure 25 ; Cartegni *et al.*, 2002 ; Pan *et al.*, 2008 ; Scotti and Swanson, 2016).

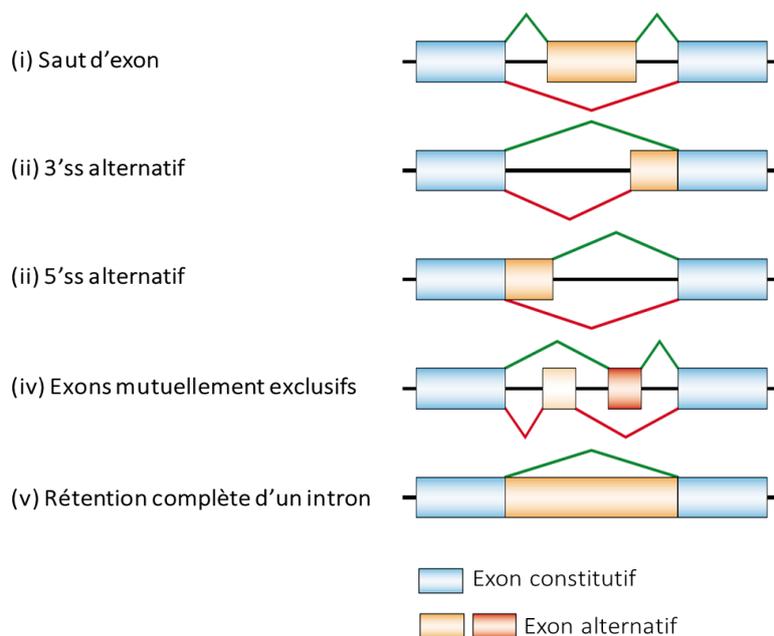


Figure 25. Les différents évènements d'épissage alternatif (adapté de Cartegni *et al.*, 2002). Dans chaque cas, la ligne verte indique l'épissage constitutif et la ligne rouge illustre l'épissage alternatif.

III.4.1. La régulation de l'épissage alternatif

L'épissage alternatif est possible grâce à la modulation de la reconnaissance des motifs d'épissage par le spliceosome. Ce mécanisme est influencé par plusieurs facteurs : la force des sites d'épissage, la présence d'éléments de régulation de l'épissage, le niveau d'expression des protéines interagissant avec le spliceosome ainsi que la structure secondaire des ARNm (Pan *et al.*, 2008 ; Baralle and Giudice, 2017). Les sites d'épissage classiquement considérés comme « forts » sont généralement présents au niveau des exons constitutifs tandis que les sites d'épissage dits « faibles » sont plutôt identifiés au niveau des exons alternatifs (Fu and Ares, 2014). La reconnaissance de ces sites par la machinerie d'épissage dépend de la séquence spécifique de ce site mais également de l'existence de séquences régulatrices avoisinantes. En effet, les séquences *cis* régulatrices de l'épissage, qu'elles soient exoniques ou introniques, en interagissant avec des protéines régulatrices, vont permettre d'influencer l'utilisation des sites d'épissage, en réprimant celle des sites d'épissage constitutifs ou, à l'inverse, en favorisant celle des sites d'épissage alternatifs en fonction de la nature inhibitrice ou activatrice de l'élément considéré, respectivement (Wang and Burge, 2008). Il est actuellement admis que les éléments *cis* régulateurs de type activateurs sont plus largement concentrés au niveau des exons constitutifs, tandis que les éléments de type inhibiteurs sont majoritairement retrouvés au niveau des introns et sont donc généralement associés au contrôle de l'épissage alternatif (Barash *et al.*, 2010). En outre, les éléments régulateurs de l'épissage, qu'ils soient exoniques ou introniques, présentent une conservation plus élevée dans et autour des exons alternatifs comparativement aux exons constitutifs, soulignant ainsi l'importance de ces éléments dans la définition des exons alternatifs (Wang and Burge, 2008 ; Coelho and Smith, 2014).

L'épissage alternatif peut également être modulé par les facteurs régulateurs de l'épissage dont la nature, le niveau d'expression ou la localisation cellulaire varient selon le tissu, le stade de développement mais également en réponse à des stimuli (Pan *et al.*, 2008 ; Baralle and Giudice, 2017). Un exemple emblématique est celui de la protéine NOVA, un facteur d'épissage neuronal exprimé spécifiquement dans le tissu cérébral et qui est à l'origine de la régulation d'un réseau de près de 700 transcrits alternatifs dont les fonctions interviennent dans le développement synaptique (Buckanovich *et al.*, 1993 ; Zaghlool *et al.*, 2014).

Enfin, les structures secondaires de l'ARN peuvent engendrer des topologies particulières de la séquence du pré-ARNm, notamment des structures en épingle à cheveux et ainsi, masquer des signaux d'épissage. Selon la nature inhibitrice ou activatrice de l'élément masqué, ces structures secondaires peuvent réprimer ou favoriser l'épissage alternatif (Shepard and Hertel, 2008 ; Coelho and

Smith, 2014). On peut notamment citer l'exemple de la structure en épingle à cheveux localisée au niveau du 5'ss de l'exon 10 du gène *MAPT* qui permet de réguler l'épissage alternatif de cet exon afin de produire deux protéines tau contenant 3 (3R-tau) ou 4 (4R-tau) régions de liaison aux microtubules (Grover *et al.*, 1999 ; Park *et al.*, 2016). A l'état physiologique, le niveau d'expression des deux protéines est similaire. Toutefois, lorsque la structure en épingle à cheveux est déstabilisée par des variations de séquence, cela conduit à une augmentation de l'inclusion de l'exon 10 déséquilibrant ainsi les ratios protéiques en faveur de la protéine 4R-tau à l'origine de la maladie neurodégénérative FTDP-17 (*FrontoTemporal Dementia and Parkinsonism linked to chromosome 17*) (Grover *et al.*, 1999 ; Park *et al.*, 2016).

Au cours de ces dernières années, l'implémentation des technologies de séquençage à haut débit dans le cadre de l'analyse du transcriptome (RNA-seq) a permis de grandes avancées dans la caractérisation des différentes isoformes issues d'un épissage alternatif. Pour autant, notre connaissance de la régulation de l'épissage alternatif reste encore parcellaire car dépendante d'un code (*splicing code*) particulièrement complexe à décrypter. En effet, les mêmes éléments régulateurs de l'épissage peuvent être reconnus par différents partenaires protéiques selon le contexte nucléotidique du pré-ARNm ou le contexte cellulaire. De plus, les facteurs protéiques peuvent être associés à des effets inhibiteurs ou activateurs en fonction du contexte dans lequel ils sont exprimés. Enfin, tous les signaux d'épissage ne sont pas encore bien définis et tous les partenaires protéiques pouvant potentiellement moduler l'épissage alternatif ne sont pas encore bien identifiés ou caractérisés. Il reste encore aujourd'hui difficile d'appréhender avec précision les mécanismes de régulation de l'épissage alternatif et de déterminer les fonctions associées aux différentes isoformes produites (Fu and Ares, 2014 ; Baralle and Giudice, 2017).

III.4.2. Le rôle de l'épissage alternatif dans la régulation de l'expression génique

L'épissage alternatif participe à la régulation de l'expression génique, notamment en regard de la production de protéines dont les rôles fonctionnels peuvent être différents voir antagonistes. Par exemple, l'isoforme de référence issue de l'épissage constitutif du gène *BCLX* conduit à la production d'une protéine qui contribue à l'inhibition de l'apoptose, tandis qu'une isoforme issue d'un épissage alternatif de l'exon 2 de ce même gène génère une protéine impliquée dans l'activation de l'apoptose (Stevens and Oltean, 2019). L'épissage alternatif peut être finement régulé en fonction du tissu, du stade de développement ou encore en réponse à un signal donné (Baralle and Giudice, 2017). Le gène *FN1* est l'un des premiers gènes à avoir bien illustré la contribution de l'épissage alternatif dans la régulation de l'expression génique tissu-dépendante (Kornblihtt *et al.*, 1984). Dans les fibroblastes,

l'épissage du pré-ARNm *FN1* induit l'inclusion d'un exon associé à un domaine fonctionnel à l'origine de la localisation tissulaire de la fibronectine générant ainsi une protéine localisée dans la matrice cellulaire, tandis que dans le foie, cet exon n'est pas inclus et est alors responsable de la production d'une isoforme soluble de la protéine qui sera sécrétée dans le plasma.

Tous les transcrits engendrés par l'épissage alternatif ne conduisent pas systématiquement à l'expansion du protéome. En effet, des écarts significatifs entre des données transcriptomiques et protéomiques ont été constatés, suggérant fortement que beaucoup d'isoformes issues de l'épissage alternatif ne sont pas à l'origine d'une protéine (Tress *et al.*, 2017). Les modifications apportées aux pré-ARNm par l'épissage alternatif, telles que des sauts d'exon, peuvent être à l'origine d'un décalage du cadre de lecture responsable de l'introduction d'un codon stop prématuré (PTC, *Premature Termination Codon*) dans les molécules d'ARNm mature (Lejeune and Maquat, 2003 ; McGlincy and Smith, 2008). Ces transcrits porteurs d'un PTC peuvent être la cible d'un système de dégradation spécifique appelé *nonsense mediated decay* (NMD) (Chang *et al.*, 2007 ; Schweingruber *et al.*, 2013).

III.4.3. Le système de surveillance des ARNm : *nonsense mediated decay* ou NMD

Le mécanisme de surveillance des ARNm, ou NMD, permet la dégradation spécifique des transcrits porteurs d'un PTC évitant ainsi la production de certaines protéines tronquées susceptibles d'inhiber la fonction de la protéine pleine longueur co-exprimée par l'allèle sauvage, du fait d'une activité de dominant négatif. Généralement, le NMD provoque la dégradation de transcrits porteurs d'un PTC situé à plus de 50 à 55 nucléotides de la dernière jonction exon/exon (Figure 26 ; McGlincy and Smith, 2008 ; Kurosaki and Maquat, 2016 ; Kurosaki *et al.*, 2019). Lors du processus d'épissage, les complexes protéiques nommés EJC (*Exon-Junction Complex*) se fixent sur l'ARN à environ 20 à 24 nucléotides en amont de chaque jonction exon/exon. Après l'export de l'ARNm vers le cytoplasme et lors du premier cycle de traduction, le ribosome provoque l'éjection des EJC au fur et à mesure de son passage et termine la traduction au niveau du codon stop naturel avant de ré-initier de nouveaux cycles de traduction. Si un PTC est présent à plus de 50 à 55 nucléotides de la dernière jonction exon/exon, le ribosome ne peut pas dissocier le(s) dernier(s) EJC présents sur l'ARNm car la traduction s'arrête de façon abrupte au niveau du PTC en formant un complexe de terminaison. L'arrêt du ribosome conduit alors au recrutement de facteurs eRFs (*eukaryotic Release Factors*) et de la protéine UPF1 (*Up-Frameshift Protein 1*). L'interaction de UPF1 avec les deux protéines UPF2 et UPF3, déjà présentes dans le complexe EJC positionné en aval, provoque alors la dégradation de l'ARNm porteur du PTC (Figure 26 ; McGlincy and Smith, 2008 ; Kurosaki and Maquat, 2016 ; Kurosaki *et al.*, 2019).

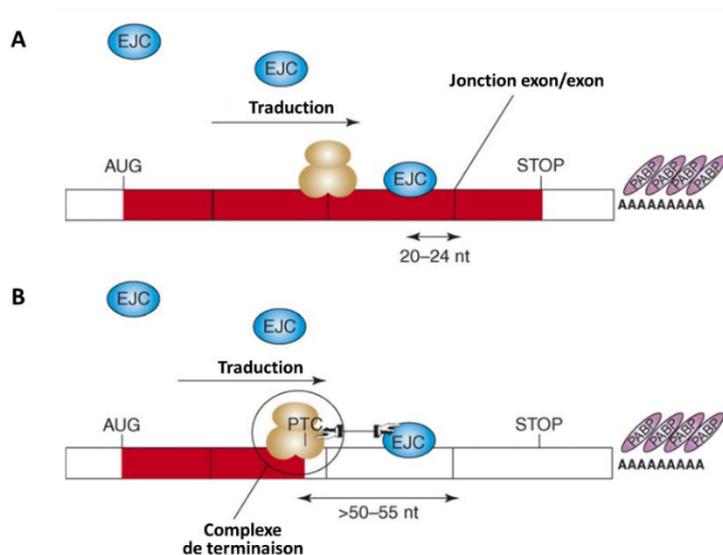


Figure 26. Le mécanisme de dégradation des transcrits porteurs d'un PTC par le système NMD (d'après McGlincy and Smith, 2008). A. Transcrit constitutif normal. B. Transcrit porteur d'un PTC cible du NMD.

Cependant, il est important de noter que près de 50% des transcrits porteurs d'un PTC échappent au NMD, pourcentage qui varie selon le gène considéré (Lindeboom *et al.*, 2019 ; Supek *et al.*, 2020). Il apparaît donc que certains mécanismes de régulation du NMD restent encore à élucider afin de déterminer avec précision quels transcrits sont la cible privilégiée de ce système de dégradation, avec de réelles implications sur l'interprétation des conséquences fonctionnelles des variations à l'origine de l'introduction d'un PTC dans un contexte de maladie monogénique.

De plus, le NMD ne constitue pas uniquement un système de surveillance de la qualité des ARNm. Il participe également à la régulation des niveaux d'expression de certains gènes, y compris de gènes codant pour des facteurs eux-mêmes impliqués dans l'épissage. Il s'agit dans ce cas, d'un mécanisme de rétrocontrôle négatif (autorégulation appelé parfois régulation croisée) via un épissage alternatif non-productif (Ni *et al.*, 2007 ; Lim *et al.*, 2020). C'est notamment le cas de certaines protéines SR qui, en se liant spécifiquement à leur pré-ARNm favorisent par exemple l'inclusion d'exons « poisons » (porteurs de PTC, donc cibles de NMD) et contrôlent ainsi directement leur propre niveau d'expression (Lareau *et al.*, 2007 ; Lareau and Brenner, 2015).

III.4.4. L'épissage alternatif des gènes *BRCA* et *MMR*

Concernant les gènes de prédisposition *BRCA* et *MMR*, de récentes études ont pu établir des catalogues exhaustifs de tous les événements d'épissage alternatifs identifiés pour chacun des gènes à partir du matériel biologique (tissus, sang, lignées lymphoblastoïdes) d'individus sains. Ces

informations sont indispensables à l'interprétation des altérations de l'épissage observées dans le contexte du syndrome sein-ovaire et du syndrome de Lynch, respectivement.

Des travaux menés à partir de matériels biologiques dérivés de sang d'individus témoins ont permis de détecter expérimentalement par RT-PCR, en plus du transcrit de référence, la présence de 68 et 27 transcrits alternatifs générés à partir des gènes *BRCA1* et *BRCA2*, respectivement, (Colombo *et al.*, 2014 ; Romero *et al.*, 2015 ; Fackenthal *et al.*, 2016) dont certaines ont été confirmées et quantifiées par la technique de RNA-seq ciblé (Figure 27 ; Davy *et al.* 2017). Ces isoformes issues d'un épissage alternatif sont le résultat de différents évènements tels que (i) des sauts d'un exon, (ii) des sauts de plusieurs exons, (iii) l'utilisation de sites donneurs et accepteurs d'épissage alternatifs conduisant à des rétentions introniques ou des délétions exoniques partielles, (iv) l'utilisation d'un codon d'initiation ou de terminaison alternatif et enfin (v) des intronisations. Il est à noter que les deux derniers évènements (iv et v) ont uniquement été rapportés dans le cas de transcrits alternatifs du gène *BRCA1* (Colombo *et al.*, 2014). Également, des combinaisons de plusieurs évènements d'épissage ont été identifiés tels que (i) des sauts d'exon concomitants à l'utilisation d'un site d'épissage alternatif ou encore (ii) l'utilisation de plusieurs sites alternatifs d'épissage pouvant éventuellement conduire à une exonisation d'une séquence intronique. Le type d'évènement d'épissage alternatif le plus fréquemment détecté dans les gènes *BRCA* reste cependant le saut d'un ou de plusieurs exons. De façon intéressante, la grande majorité des transcrits alternatifs détectés dans le sang sont également identifiés dans le tissu mammaire sain et aucun transcrit *BRCA* spécifique du tissu mammaire n'a été détecté. De plus, l'ensemble des transcrits alternatifs du gène *BRCA2* ont tous été identifiés dans au moins une des six lignées cellulaires dérivées de cellules mammaires analysées (MCF7, HCC1937, BT20, MCF10A, 184A1, 184B5) (Fackenthal *et al.*, 2016). De plus, la majorité des transcrits alternatifs identifiés comme prédominants, trois dans le gène *BRCA1* ($\Delta 8p$, $\Delta 9-10$, $\Delta 14p$) et un dans le gène *BRCA2* ($\Delta 12$), sont associés à une expression similaire dans les tissus sanguins et mammaires.

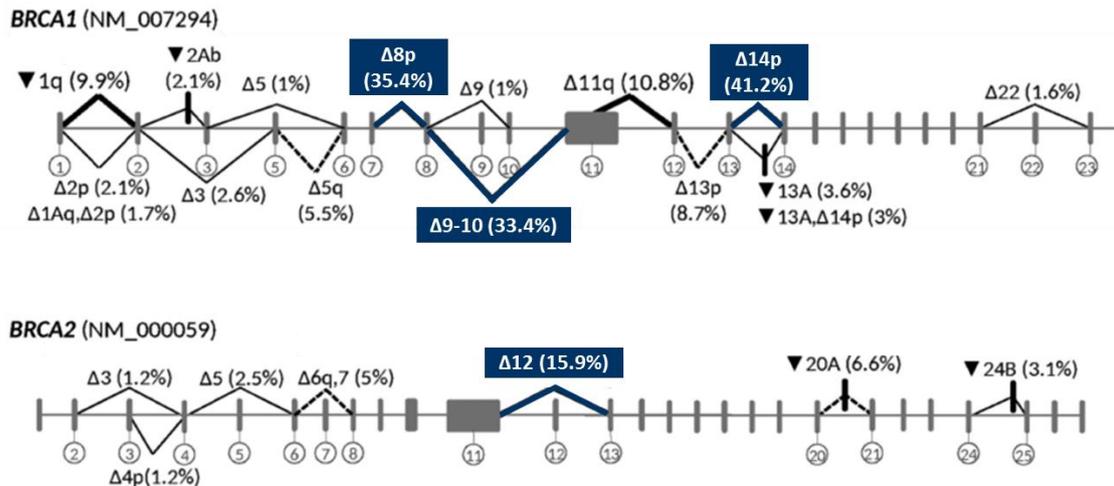


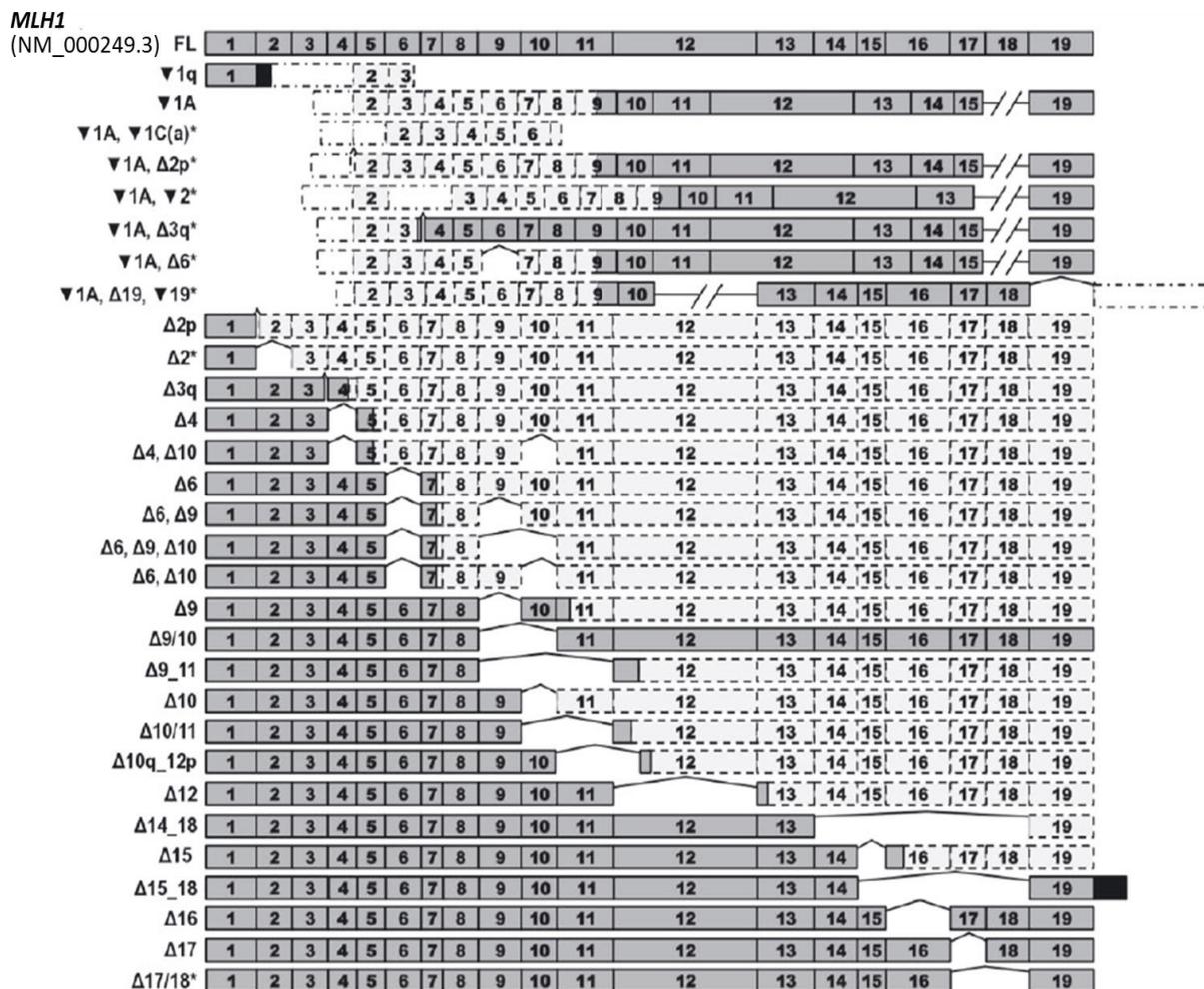
Figure 27. Représentation schématique des principaux évènements d'épissage alternatif dans les gènes *BRCA1* et *BRCA2* (d'après Davy *et al.*, 2017). ▼, inclusion d'une séquence intronique ; Δ, délétion partielle ou totale d'exon ; p, utilisation d'un site 3' d'épissage alternatif ; q, utilisation d'un site 5' d'épissage alternatif. Les cadres bleus correspondent aux évènements d'épissage alternatif prédominants ($\geq 15\%$), tous à l'origine de délétions en phase.

Bien que l'identification des transcrits alternatifs des gènes *BRCA* constitue une avancée majeure en soit, il n'en reste pas moins que la fonctionnalité et la pertinence biologiques de telles isoformes sont encore très peu appréhendées de nos jours. Si la majorité des transcrits alternatifs des gènes *BRCA* induisent un décalage du cadre de lecture conduisant à l'introduction d'un codon stop prématuré potentiellement cible du NMD ou sont non-codants à la suite de l'élimination du codon d'initiation de la traduction, certains transcrits pourraient générer des isoformes fonctionnelles. C'est notamment le cas des transcrits présentant des délétions en phase des modifications des régions terminales non codantes ou encore, des codons stops prématurés non cibles du NMD car localisés au niveau du dernier exon. Pour autant, quelques études ont déjà démontré que certains de ces transcrits sont à l'origine de protéines non fonctionnelles telles que la protéine *BRCA1* délétée pour l'exon 14 et 15 ou encore, la protéine *BRCA2* délétée pour l'exon 3 (Sevcik *et al.*, 2012 ; Caputo *et al.*, 2018). A l'inverse, certains travaux tendent à démontrer le caractère fonctionnel, au moins en partie, de quelques isoformes dont notamment celles résultant du saut de l'exon 9 et 10 de *BRCA1*, du saut des exons 4 à 7 de *BRCA2* ou encore, du saut de l'exon 12 de *BRCA2* (de la Hoya *et al.*, 2016 ; Li *et al.*, 2009 ; Thirthagiri *et al.*, 2016 ; Mesman *et al.*, 2020). Néanmoins, la fonctionnalité de la plupart des transcrits alternatifs potentiellement codants reste encore à déterminer.

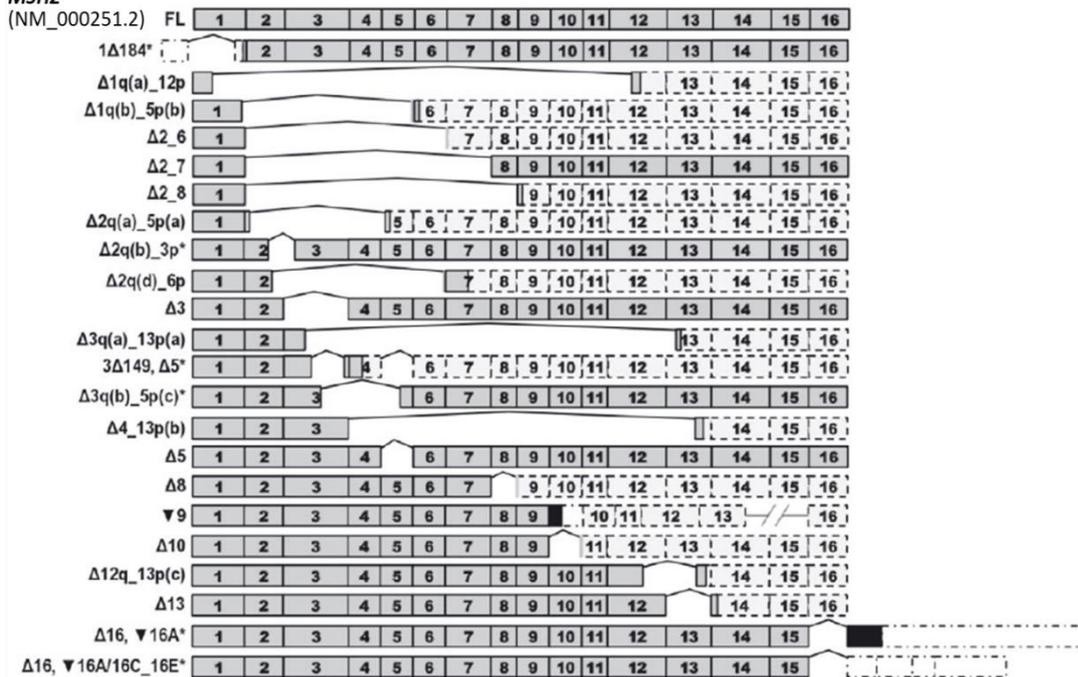
Pour les gènes *MLH1*, *MSH2*, *MSH6* et *PMS2*, ce sont 30, 22, 4 et 9 transcrits alternatifs, respectivement, qui ont été rapportés en plus du transcrit de référence après la compilation de toutes les informations disponibles dans la littérature ou dans des bases de données (Figure 28 ; Thompson

et al., 2015). Au même titre que pour les gènes *BRCA*, les épissages alternatifs des gènes *MMR* résultent de différents évènements ou combinaison d'évènements dont les plus fréquents dans le gène *MLH1* sont les sauts d'exon et dans les gènes *MSH2* et *MSH6*, l'utilisation de sites d'épissage alternatifs. Toutefois, à l'inverse des gènes *BRCA*, certains des transcrits alternatifs des gènes *MMR* ne sont pas toujours détectés selon le tissu considéré. En particulier, des transcrits alternatifs du gène *MLH1* ont été identifiés à partir de lymphocytes mais certains n'ont pas été détectés dans le tissu colorectal (Plaschke et al., 1999). Par ailleurs, seules quelques études se sont attachées à déterminer les niveaux d'expression des transcrits alternatifs des gènes *MMR* à partir de matériel biologique d'individus contrôles, représentant ainsi une limite importante à l'analyse de ces isoformes (Charbonnier et al., 1995 ; Palmirotta et al., 1998 ; Takahashi and Nagai, 2009).

Figure 28. Représentation schématique des évènements d'épissage alternatif dans les gènes *MLH1*, *MSH2*, *MSH6* et *PMS2* (d'après Thompson et al., 2015). Les rectangles gris correspondent à la séquence codante tandis que les rectangles en pointillés représentent la séquence non codante. FL (*Full-Lenght*) transcrit pleine longueur ; ▼, inclusion d'une séquence intronique ; Δ, délétion partielle ou totale d'exon ; p, utilisation d'un site 3' d'épissage alternatif ; q, utilisation d'un site 5' d'épissage alternatif.



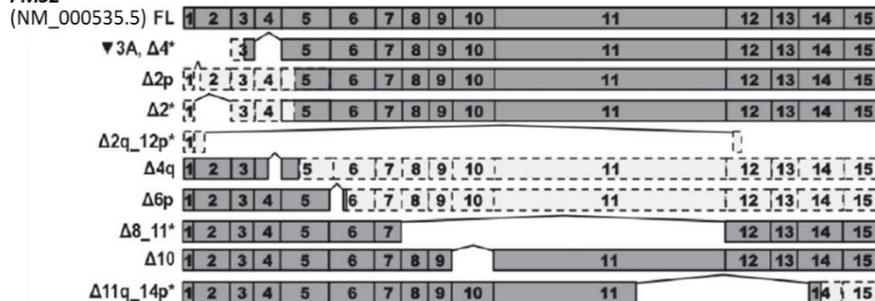
MSH2



MSH6



PMS2



Comme pour les gènes *BRCA*, il est attendu que la plupart des transcrits alternatifs des gènes *MMR* soient non fonctionnels car ils sont soit non-codants, soit responsables de l'introduction d'un PTC et donc potentiellement cibles du NMD ou bien encore, à l'origine de la production de protéines tronquées ou de protéines associées à un effet dominant-négatif (Thompson *et al.*, 2015). Aujourd'hui, seulement trois isoformes protéiques issues de transcrits alternatifs du gène *MLH1* ($\Delta 9$ -10, $\Delta 16$, $\Delta 17$) ont fait l'objet d'investigations expérimentales (Trojan *et al.*, 2002 ; Peasland *et al.*, 2010). Une perte d'activité MMR a été rapportée pour ces trois isoformes du fait d'un défaut d'interaction soit avec l'hétérodimère MutS α (pour *MLH1* $\Delta 9$ -10), soit avec le partenaire protéique PMS2 (pour *MLH1* $\Delta 16$ et

Δ17). Par ailleurs, une étude suggère que l'isoforme protéique correspondant au transcrit alternatif sans les exons 9 et 10 pourrait exercer un effet dominant négatif (Peasland *et al.*, 2010).

Les études de caractérisation fonctionnelle des isoformes alternatives des gènes *BRCA* et *MMR* restent encore trop peu nombreuses et soulèvent, à raison, la question du rôle biologique de ces transcrits. Une meilleure compréhension de l'importance de l'épissage alternatif de ces gènes est primordiale pour l'interprétation des défauts d'épissage, identifiés dans le cadre du syndrome sein-ovaire ou le syndrome de Lynch.

IV. Les défauts d'épissage induits par des variations génétiques

L'épissage est donc un processus extrêmement dynamique et particulièrement complexe étant donné les multiples signaux d'épissage (éléments *cis*) présents sur les pré-ARNm mais également du fait de la myriade de facteurs *trans* intervenant dans la régulation de ce mécanisme moléculaire. Ce processus contribue à la régulation de l'expression génique mais peut être altéré par des variations génétiques si celles-ci touchent un des nombreux acteurs intervenant dans ce processus avec pour conséquence éventuelle une modification du profil d'épissage des transcrits. Longtemps, il a été considéré que les maladies mendéliennes étaient le fait de variations responsables d'un effet direct sur la protéine, au niveau traductionnel, via notamment le gain d'un PTC (variations non-sens, délétions ou insertions à l'origine d'un décalage du cadre de lecture), la plupart de ces maladies ayant dans un premier temps été associées à ce type de variations. Cependant, et ce grâce à de très nombreuses années de recherche, il est désormais bien établi qu'une proportion importante des variations à l'origine des maladies mendéliennes correspond à des variations responsables d'une altération de l'épissage (Tableau 14 ; Lopez-Bigas *et al.*, 2005 ; Baralle *et al.*, 2009 ; Lim *et al.*, 2011). Si ces variations splicéogéniques représentent près d'un tiers des mutations causales dans l'étiologie de certaines maladies mendéliennes, c'est que théoriquement toute variation nucléotidique, qu'elle soit exonique ou intronique, est susceptible d'altérer l'épissage (Wang and Cooper, 2007 ; Ward and Cooper, 2010 ; Scotti and Swansson, 2016). Selon le type d'élément altéré et son contexte nucléotidique, différentes conséquences peuvent en résulter. De nombreux outils bioinformatiques de prédiction mais également un certain nombre d'approches expérimentales ont été implémentés dans le but de contribuer à la mise en évidence et à l'interprétation précise des variations à l'origine d'un défaut d'épissage.

Gene	Disease/phenotype	Splicing/total mutation* (%)
<i>ATM</i>	Ataxia telangiectasia	18
<i>BRCA1</i>	Breast cancer predisposition	9
<i>CADM</i>	Medium chain acyl CoA dehydrogenase deficiency	10
<i>CFTR</i>	Cystic fibrosis	14
<i>DMD</i>	Duchenne muscular dystrophy	9
<i>HBA1/2</i>	Blood disorders (thalassaemias, anaemia etc)	3
<i>HBB</i>	Blood disorders (thalassaemias, anaemia etc)	10
<i>HPRT</i>	Hypoxanthine-guanine phosphoribosyltransferase 1 deficiency	15
<i>IKBKAP</i>	Dysautonomia, familial	33
<i>MAPT</i>	Frontotemporal dementia and Parkinsonism	33
<i>MLH1</i>	Colorectal cancer	18
<i>MSH2</i>	Colorectal cancer	9
<i>NF1</i>	Neurofibromatosis type 1	19
<i>NF2</i>	Neurofibromatosis type 2	22
<i>RHO</i>	Retinitis pigmentosa	3
<i>SMN1/2</i>	Spinal muscular atrophy	4
<i>WT1</i>	Wilms tumour	11

Tableau 14. Estimation du pourcentage de variations splicéogéniques par rapport à l'ensemble des variations pathogènes à l'origine de maladies mendéliennes (d'après Baralle *et al.*, 2009). Les fréquences ont été calculées à partir de la base de données HGMD (*Human Gene Mutation Database*) en 2008.

Ce n'est que quelques années seulement après la découverte du mécanisme d'épissage qu'il a été démontré que des variations génétiques pouvaient impacter son bon fonctionnement. En 1981, l'équipe de Richard A. Flavell, biologiste moléculaire anglais, rapporte l'existence d'une variation dans le premier intron du gène de la β -globine qui, avec l' α -globine, constitue la forme d'hémoglobine la plus courante chez l'homme adulte (Busslinger *et al.*, 1981). De façon intéressante ils ont pu montrer que cette variation créait un nouveau site 3'ss dans le premier intron et était donc à l'origine de la production d'un transcrit aberrant présentant la rétention des 19 derniers nucléotides de cet intron. La présence d'une telle insertion au sein de la molécule d'ARNm mature conduit au décalage du cadre de lecture et à l'introduction d'un codon stop prématuré au niveau du 36^{ème} codon de la protéine. L'équipe de Richard A. Flavell émet alors l'hypothèse qu'une telle molécule d'ARNm conduira à la production d'une protéine tronquée non fonctionnelle (Busslinger *et al.*, 1981). Toutefois, l'équipe du scientifique américain Sherman M. Weissan, après une analyse des profils d'épissage de l'ARNm issus des réticulocytes de deux individus porteurs de la variation d'intérêt, conclura qu'une telle molécule d'ARNm est vraisemblablement instable, dégradée et/ou pas correctement transportée vers le cytoplasme (Fukumaki *et al.*, 1982). Le fait que cette variation fut identifiée dans le génome de patients atteints de β -thalassémie constitue la première preuve qu'une variation génétique peut affecter le mécanisme d'épissage et, de ce fait, être à l'origine d'une maladie mendélienne. Ces travaux ont

permis une première caractérisation d'une anomalie d'épissage et, dès lors, ont ouvert la voie à de très nombreuses études investiguant la nature des variations responsables d'altérations de l'épissage, maintenant qualifiées de splicéogéniques, mais également les différents mécanismes par lesquels les variations peuvent conduire à de telles modifications.

IV.1. Les mécanismes d'altération de l'épissage

L'émergence d'un défaut d'épissage associé à une maladie mendélienne peut présenter diverses origines. En effet, les variations génétiques peuvent directement affecter l'épissage en modifiant la séquence soit d'un élément *cis* au niveau du pré-ARNm soit d'un facteur *trans* (Tableau 15 ; Wang and Cooper, 2007 ; Ward and Cooper, 2010 ; Scotti and Swanson, 2016). La distinction entre ces différentes altérations est essentielle car selon la nature du composant altéré, les conséquences seront différentes. De fait, une modification de la séquence d'un élément *cis* a un impact direct sur l'expression d'un gène en particulier tandis qu'une altération d'un facteur *trans* peut affecter l'expression de plusieurs gènes. Il a d'ailleurs été très largement montré que la majorité des variations à l'origine d'une anomalie d'épissage associée à une maladie mendélienne sont des variations altérant un élément *cis* car celles affectant les facteurs *trans* sont souvent létales au stade embryonnaire reflétant ainsi le caractère essentiel de ces facteurs.

Maladie	Prévalence	Gènes ^a	Type d'altération	Conséquence sur la protéine
Bêta+ thalassémie	1/100 000	<i>HBB</i>	Site 3'/5' d'épissage (<i>cis</i>)	Insertion, délétion d'acides aminés ou protéine tronquée
Dystrophie musculaire de Duchenne	1/3 300 (garçons)	<i>DMD</i>	Site 3'/5' d'épissage (<i>cis</i>)	Protéine tronquée
Mucoviscidose	Entre 1/10 000 et 1/8 000	<i>CFTR</i>	Site 3'/5' d'épissage et région riche en pyrimidines (<i>cis</i>)	Insertion, délétion d'acides aminés ou protéine tronquée
Ataxie-télangiectasie	1/100 000	<i>ATM</i>	Site 3'/5' d'épissage (<i>cis</i>)	Délétion d'acides aminés ou protéine tronquée
Porphyrie érythroïdétique congénitale	< 1 / 1 000 000	<i>URO5</i>	Point de branchement (<i>cis</i>)	Protéine tronquée ?
Hémophilie B	1/30 000	<i>FIX</i>	Région riche en pyrimidines (<i>cis</i>)	Protéine tronquée ?
Démence fronto-temporale et parkinsonisme liée au chromosome 17	Entre 1/100 000 et 1/10 000	<i>MAPT</i>	Séquence régulatrice de l'épissage (<i>cis</i>)	Délétion d'acides aminés
Amyotrophie spinale	1/30 000	<i>SMN1/SMN2</i>	Séquence régulatrice de l'épissage (<i>cis</i>)	Protéine tronquée
Dystrophie myotonique de type 1	Entre 1/10 000 et 1/2 000	<i>MAPT</i>	Facteur d'épissage (<i>trans</i>)	NA
Dystrophie myotonique de type 2	1/100 000	<i>ZNF9</i>	Séquestration de facteur d'épissage	NA
Sclérose latérale amyotrophique	Entre 1/100 000 et 1/10 000	<i>C9ORF72, SOD1, TARDBP et FUS</i>	Facteur d'épissage (<i>trans</i>) et séquestration de facteur d'épissage	NA

Tableau 15. Exemples de maladies mendéliennes associées à une altération d'un élément *cis* ou d'un facteur *trans* d'épissage (d'après de la Grange, 2016)

IV.1.1. Altération des signaux *cis* d'épissage

Toutes les variations, qu'elles soient exoniques ou introniques, sont susceptibles d'altérer les signaux *cis* d'épissage et ce, quels que soient leurs effets prédits sur la protéine (y compris des variations non-sens, faux-sens ou encore synonymes). Selon le type d'élément altéré et son contexte nucléotidique, différentes conséquences au niveau du transcrit mature peuvent en résulter : un saut d'exon, une délétion d'une partie d'exon, une rétention soit d'une partie d'intron directement adjacente à l'exon soit de l'intron en entier, la création d'un nouvel exon (pseudo-exon) ou encore, une modification de l'équilibre en transcrits constitutifs/alternatifs (Figure 29 ; Ward and Cooper, 2010 ; Thompson *et al.*, 2015). L'ensemble de ces évènements peuvent potentiellement conduire à l'introduction d'un codon stop prématuré avec pour conséquence une dégradation de la molécule d'ARNm par le système NMD ou la production d'une protéine modifiée avec des fonctions altérées. Les variations ainsi capables d'affecter l'épissage constitutif comme l'épissage alternatif se distinguent en deux sous-groupes : (i) les variations altérant les principaux signaux d'épissage tels que les sites 5' et 3' d'épissage, le site de branchement et la région riche en pyrimidine et (ii) celles modifiant les éléments de régulation de l'épissage, exoniques ou introniques.

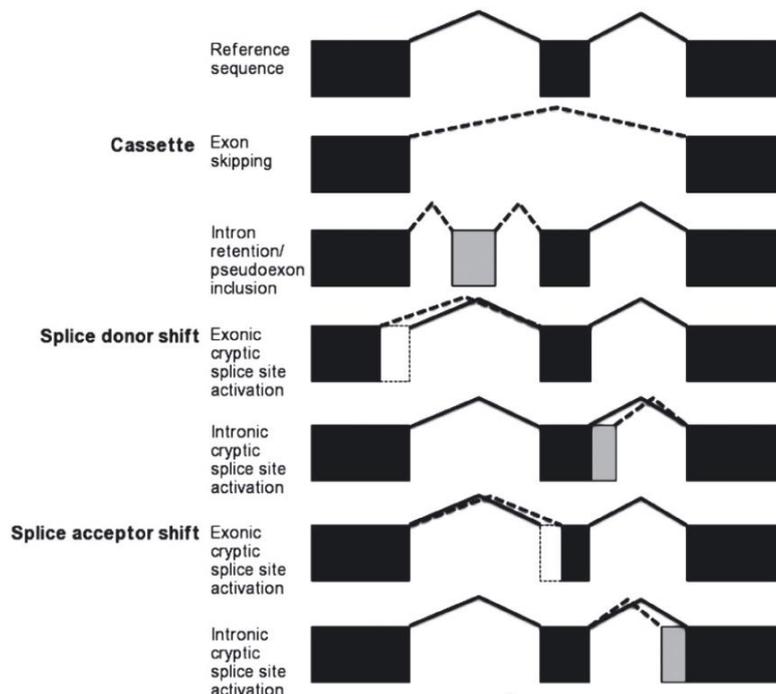


Figure 29. Les différentes conséquences des anomalies d'épissage au niveau de l'ARNm mature (d'après Thompson *et al.*, 2015). Les carrés noirs représentent les exons constitutifs, les carrés blancs délimités par des pointillés correspondent aux régions exoniques délétées et les carrés gris indiquent les régions introniques retenues et devenues exoniques. Dans chaque cas, la ligne simple illustre l'épissage constitutif et la ligne en pointillés indique l'épissage altéré.

IV.1.1.a. Altération des signaux principaux

Les variations splicéogéniques les plus fréquemment décrites à ce jour comme étant à l'origine d'une maladie mendélienne sont celles affectant les dinucléotides introniques les plus conservés des sites d'épissage, en position +1 et +2 pour le 5'ss et -1, -2 pour le 3'ss (IVS±1/2, *InterVening Sequence*) (Krawczak *et al.*, 1992 ; Krawczak *et al.*, 2007). En effet, et comme précisé dans la partie III.3.1., les dinucléotides GT et AG respectivement aux positions IVS+1/2 et IVS-1/2 sont présents dans 99% des introns et sont essentiels à la définition des jonctions exon/intron. Ainsi, quelle que soit la variation considérée, l'altération conduira, dans la majorité des cas, à une absence totale de reconnaissance du 5'ss ou 3'ss constitutif par le splicéosome. Il est toutefois important de noter que les substitutions T>C localisées en position +2 au niveau des 5'ss (GT>GC) ne sont pas toujours associées à un défaut d'épissage complet. En effet, des travaux récents ont montré que, dans 15 à 18% des cas, ce type de variations permet la production de transcrits normaux à des taux plus ou moins importants (1 à 89%) (Lin *et al.*, 2019 ; Chen *et al.*, 2020). L'épissage peut également être perturbé par toute autre variation localisée au niveau des sites d'épissage en dehors des dinucléotides introniques les plus conservés, notamment les variations localisées sur les premiers et derniers nucléotides de l'exon, mais également celles en position -3 et +3 à +6 des introns flanquants (Krawczak *et al.*, 2007 ; Grodecka *et al.*, 2014a). Selon l'anomalie considérée, l'altération résultante pourra être à l'origine d'une diminution de la reconnaissance du 5'ss ou 3'ss par le splicéosome, d'une absence totale de reconnaissance ou encore d'une augmentation de celle-ci.

Lorsque le motif d'un site d'épissage constitutif est altéré voir détruit, le splicéosome ne pourra reconnaître ce site comme tel et cherchera donc à utiliser un autre site d'épissage accessible, en général le plus proche (Krawczak *et al.*, 2007, Ward and Cooper, 2010). Selon la position de ce site, diverses conséquences sont possibles. Dans le cas où le site d'épissage le plus proche est le site d'épissage constitutif de l'exon voisin, cela entraînera un saut d'exon (Figure 29). Par exemple, la variation c.884+4A>G localisée dans l'intron 10 du gène *MLH1* et identifiée dans le génome d'un patient atteint d'un syndrome de Lynch, altère le 5'ss de l'exon 10 et provoque ainsi le saut de cet exon (Rahner *et al.*, 2007). Cette anomalie d'épissage induit un décalage du cadre de lecture et conduit de ce fait à l'introduction d'un codon stop prématuré. Le transcrit résultant du saut de l'exon 10 pourra alors être dégradé par le système NMD, comme le suggère une étude menée par notre équipe (Soukariéh *et al.*, 2016). Si la molécule d'ARNm mature n'est pas complètement dégradée, elle conduira alors à la production d'une protéine tronquée non fonctionnelle. Dans certains cas, le saut d'exon peut être en phase et ne pas introduire de PTC, la molécule d'ARNm ne fera alors pas l'objet d'une dégradation et une protéine dépourvue d'un fragment interne sera produite. C'est par exemple

le cas de la variation c.8487+1G>A localisée dans l'intron 19 du gène *BRCA2* et qui a été identifiée au sein de plusieurs familles atteintes du syndrome sein-ovaire (Agata *et al.*, 2003). L'exon 19 est en phase mais code pour un domaine protéique essentiel au bon fonctionnement de la protéine. La variation c.8487+1G>A en altérant le 5'ss de l'exon 19 induit le saut de cet exon et ainsi, la perte de fonction de la protéine.

Le splicéosome peut également reconnaître un site d'épissage cryptique à proximité du site d'épissage altéré, générant ainsi d'autres types d'évènements tels que la rétention d'une partie d'intron ou la délétion d'une partie d'exon (Figure 29). Le paysage génomique est non seulement ponctué de sites d'épissage qui définissent les exons constitutifs mais également de sites d'épissage alternatifs qui sont responsables de la production de transcrits alternatifs ainsi que de sites d'épissage cryptiques, *i.e.* des sites d'épissage non reconnus dans un contexte naturel par le splicéosome pour la production des transcrits constitutifs ou alternatifs. Si le site d'épissage alternatif ou cryptique le plus proche du site altéré est localisé dans l'exon, son utilisation conduira à une délétion d'une partie de l'exon altéré (Figure 29). C'est par exemple le cas de la variation *BRCA1* c.302-2del identifiée chez des patients atteints du syndrome sein-ovaire et responsable de la destruction du 3'ss constitutif et de l'activation concomitante d'un 3'ss cryptique situé 10 nucléotides en aval, au sein de l'exon (Southey *et al.*, 2003 ; Chen *et al.*, 2006). Cette variation entraîne ainsi la délétion des 10 premiers nucléotides de l'exon 7 de *BRCA1* aboutissant à un décalage du cadre de lecture et l'introduction d'un PTC dans le transcrit mature. En revanche, si le site d'épissage alternatif ou cryptique le plus proche du site constitutif altéré est localisé dans un intron flanquant l'exon, son utilisation induira une rétention d'une partie de l'intron (Figure 29). Par exemple, la variation *BRCA1* c.4986+6T>G retrouvée chez des individus atteints du syndrome sein-ovaire, conduit à l'activation d'un 5'ss cryptique à 65 nucléotides en aval du 5'ss constitutif de l'exon 16 à la suite de l'altération de ce dernier (Chen *et al.*, 2006). Cette variation conduit donc à la rétention de 65 nucléotides introniques et induit ainsi un décalage du cadre de lecture et l'introduction d'un PTC.

Les variations à distance des sites d'épissage constitutifs peuvent impacter l'épissage sans modifier directement le site naturel du fait de la création d'un site d'épissage *de novo* ou via l'augmentation de la force d'un site d'épissage alternatif ou cryptique (Figure 29). Il a également été rapporté le cas de variations n'altérant pas les sites d'épissage constitutifs mais conduisant à l'introduction d'un nouveau site d'épissage dans une région intronique, susceptible de créer un nouvel exon appelé pseudo-exon si un site complémentaire préexiste à proximité (Figure 29). En 1995, il a ainsi été montré, pour la première fois dans le gène *CFTR*, que la variation c.1680-886A>G localisée dans l'intron 11 conduisait à la création d'un nouveau 5'ss qui a pour conséquence l'insertion d'un

pseudo-exon hors phase, nommé exon 11b, dans la molécule d'ARNm mature (Chillon *et al.*, 1995). Il est très bien documenté que la protéine CFTR est un canal ionique dont le dysfonctionnement est à l'origine de la mucoviscidose, la maladie génétique à transmission récessive la plus fréquente dans la population caucasienne. Désormais, l'inclusion de pseudo-exons est un mécanisme bien connu de l'étiologie des maladies mendéliennes. En 2012, il a notamment été rapporté la première variation intronique profonde à l'origine de l'introduction d'un pseudo-exon dans les gènes *BRCA* (Anczukow *et al.*, 2012). En effet, la variation *BRCA2* c.6937+594T>G a été identifiée au sein de 9 familles françaises évocatrices d'un syndrome sein-ovaire mais dont aucune variation pathogène n'avait été identifiée dans la séquence codante des gènes *BRCA*. Il est apparu que cette variation intronique profonde induisait une augmentation de la force d'un 5'ss cryptique et conduisait ainsi à l'inclusion d'un pseudo-exon de 95 nucléotides entre l'exon 12 et 13 du gène *BRCA2*, responsable d'un décalage du cadre de lecture. Depuis cette étude, la pathogénicité de cette variation a été questionnée du fait de sa fréquence élevée dans la population générale (Dutil *et al.*, 2018). Des travaux récents ont permis l'identification de 2 nouvelles variations à l'origine de pseudo-exons dans le gène *BRCA1*, correspondant à une variation ponctuelle intronique profonde (c.4185+4105C>T) dans l'intron 12 (Montalban *et al.*, 2019) mais également à l'insertion d'un rétrotransposon de type SINE (*Short Interspersed Nuclear Element*)-VNTR (*Variable Number Tandem Repeat*)-Alu dans l'intron 13 (Walsh *et al.*, 2020). Ces études soulignent ainsi l'importance de séquencer, dans leur intégralité, les exons mais également les introns des gènes, dans la mesure du possible.

En dehors des sites d'épissage, le site de branchement et la région riche en pyrimidine peuvent également faire l'objet d'altérations conduisant à un impact sur l'épissage. En effet, ces signaux sont essentiels dans l'interaction du spliceosome et des facteurs de régulation avec la séquence du pré-ARNm. Toutefois, la caractérisation expérimentale de ces régions reste encore incomplète et la dégénérescence de ces motifs rend difficile l'identification d'une variation potentiellement spliceogénique dans ces signaux (Abramowicz and Gos, 2018). Cependant, il apparaît que des variations localisées au niveau de la séquence du site de branchement puissent conduire à un saut d'exon ou encore à une rétention d'intron (Caminsky *et al.*, 2014). C'est notamment le cas de la variation *XPC* c.413-24A>G identifiée chez des patients atteints de xeroderma pigmentosum, une maladie génétique héréditaire rare responsable d'une sensibilité extrême aux rayons ultraviolets du soleil. En effet, cette variation en altérant le site de branchement ne permet pas l'interaction du snRNP U2 avec la molécule du pré-ARNm et conduit donc au saut hors phase de l'exon 4 du gène *XPC* (Khan *et al.*, 2010). Il est intéressant de noter qu'environ 3% de transcrits pleine longueur sont malgré tout exprimés du fait de l'utilisation d'un site de branchement alternatif très faible en position -9 dans l'intron 3 (Khan *et al.*, 2010). Jusqu'en 2013, moins de vingt variations avaient été décrites comme

étant capables d'altérer la séquence du point de branchement (Lewandowska, 2013). Depuis, plusieurs études ont rapporté de nouvelles variations conduisant à l'altération de sites de branchement, notamment dans les gènes *BRCA1*, *MLH1* ou encore *RAD51C* (Houdayer *et al.*, 2008 ; Wappenschmidt *et al.*, 2012 ; Leman *et al.*, 2020). Cependant, les variations altérant le site de branchement ou la région riche en pyrimidine restent tout de même très rarement identifiées et de ce fait, elles ne font pas systématiquement l'objet d'une recherche dans le cadre du diagnostic mais uniquement lorsqu'aucune autre variation n'a été identifiée par séquençage.

IV.1.1.b. Altération des signaux auxiliaires

Tout comme les sites d'épissage, le site de branchement et la région riche en pyrimidine, les séquences régulatrices de l'épissage sont également susceptibles d'être altérées par des variations nucléotidiques. Il a notamment été estimé que près d'un quart des variations faux-sens sont à l'origine d'un défaut d'épissage du fait de la création d'un élément de régulation inhibiteur de l'épissage (ESS) et/ou via la destruction d'un élément de régulation activateur de l'épissage (ESE) (Liu *et al.*, 2001 ; Sterne-Weiler *et al.*, 2011). Le modèle le plus emblématique de maladie mendélienne provoquée par une variation synonyme à l'origine d'un défaut d'épissage est celui de l'amyotrophie spinale (SMA, *Spinal Muscular Atrophy*), une maladie neuromusculaire caractérisée par un défaut en protéines SMN (*Survival Motor Neuron*) conduisant à une dégénérescence progressive des motoneurones (Cartegni *et al.*, 2006). Chez l'homme, la protéine SMN est codée par le gène *SMN1* ainsi que le gène *SMN2*, résultat d'une duplication génique de *SMN1* différant de seulement 5 nucléotides de la copie originale. Tandis que le gène *SMN1* permet la production de la protéine SMN pleine longueur, le gène *SMN2* conduit majoritairement à la production d'une protéine tronquée, rapidement dégradée. Cette production anormale est le résultat direct d'un des nucléotides différant de *SMN1*. En effet, dans le gène *SMN2*, une substitution sur le 6^{ème} nucléotide d'une cytosine vers une thymine (c.840C>T, p.phe280=) est responsable du saut de l'exon 7 dans la majorité des transcrits produits par *SMN2*, résultant ainsi en une protéine non fonctionnelle. Deux modèles ont été proposés afin d'expliquer comment une telle variation exonique pouvait agir directement sur l'épissage. Tout d'abord, il a été montré que la variation c.840C>T était localisée dans une séquence ESE normalement ciblée par la protéine SR SF2/ASF, un facteur de régulation activateur de l'épissage (Cartegni and Krainer, 2002). Cette substitution conduit ainsi à la perturbation du motif ESE avec pour conséquence l'absence de fixation du facteur SF2/ASF sur la molécule du pré-ARNm. Peu de temps après ces travaux, une seconde étude a démontré que cette variation altère non seulement une séquence ESE mais est également responsable de la création d'un motif ESS sur lequel se fixe la protéine hnRNPA1, favorisant ainsi l'exclusion de l'exon 7 (Kashima and Manley, 2003). En conséquence, si le gène *SMN1* est muté, le gène

SMN2 ne sera pas en capacité de compenser la déficience en protéines SMN, résultant ainsi en une amyotrophie spinale. Les deux modèles décrits sont non exclusifs et illustrent très bien le délicat équilibre entre les différents éléments activateurs et inhibiteurs de l'épissage (Cartegni *et al.*, 2006).

Ainsi, un seul changement nucléotidique peut détruire un élément de régulation de l'épissage mais peut également faire basculer cet élément d'activateur à inhibiteur et vice versa. Toutefois, ces éléments n'étant pas encore entièrement caractérisés, relativement peu de variations associées à des maladies mendéliennes ont été décrites comme à l'origine de leur altération. Pourtant, des analyses mutationnelles ciblées sur des exons modèles ont récemment révélées qu'une proportion importante des variations ponctuelles identifiées dans ces exons conduisent à un défaut d'épissage. Ainsi, 79% des variations ponctuelles analysées dans l'exon 7 du gène *SMN2* sont responsables d'une altération de l'épissage (Cartegni *et al.*, 2006 ; Singh *et al.*, 2007). Notre équipe a démontré que le nombre de variations à l'origine d'une modification des éléments ESR étaient vraisemblablement sous-estimé, notamment dans les gènes de prédisposition aux cancers (Di Giacomo *et al.*, 2013 ; Soukarieh *et al.*, 2016). En prenant comme modèles d'étude l'exon 7 de *BRCA2* (Di Giacomo *et al.*, 2013) ou encore l'exon 10 de *MLH1* (Soukarieh *et al.*, 2016), une analyse systématique des variations répertoriées dans ces exons a permis de révéler qu'un nombre important de variations (jusqu'à 66%, *MLH1* exon 10) étaient à l'origine d'une modification de l'épissage via une altération des ESR.

Au-delà des exons, les variations peuvent également altérer les séquences régulatrices introniques. C'est par exemple le cas d'une délétion de 4 nucléotides dans l'intron 6b du gène *CFTR* (c.1002-1110_1002-1113del) (Faa *et al.*, 2009). De façon intéressante, les patients porteurs de cette altération présentent des transcrits avec une séquence supplémentaire de 101 nucléotides par rapport au transcrit pleine longueur. Ces transcrits sont le résultat de l'insertion d'un exon cryptique dans l'intron 6b du gène *CFTR* qui n'est normalement pas reconnu par le splicéosome et donc, non incorporé dans la molécule d'ARNm mature. En effet, la délétion de 4 nucléotides est à l'origine de la destruction d'un ISS normalement reconnu par la protéine hnRNP A2/B1 empêchant ainsi l'insertion de l'exon cryptique et cette altération est également responsable de la création d'un ISE permettant au facteur activateur de l'épissage SRp75 de médier la reconnaissance de l'exon cryptique par le splicéosome (Faa *et al.*, 2009). Du fait de la destruction d'un élément inhibiteur et de la création d'un élément activateur, la variation c.1002-1110_1002-1113del est donc à l'origine de l'insertion d'un exon cryptique qui induit un décalage du cadre de lecture. Aujourd'hui, le nombre de variations à l'origine d'une modification des ISR reste restreint, vraisemblablement du fait de notre incapacité à appréhender l'impact des variations introniques sur de tels motifs encore très mal définis.

IV.1.2. Altération des facteurs *trans* d'épissage

Bien que les variations affectant les éléments *cis* représentent la majorité des variations splicéogéniques identifiées comme étant à l'origine d'une maladie mendélienne, d'autres variations sont susceptibles d'altérer l'épissage en modifiant les facteurs *trans* (Tableau 16). Contrairement au nombre de variations toujours plus élevé touchant les éléments *cis*, les variations décrites comme capables d'altérer les composants ubiquitaires du splicéosome et les facteurs régulateurs d'épissage sont relativement peu nombreuses. Pour cause, les anomalies des facteurs *trans* peuvent conduire au dérèglement du processus d'épissage de plusieurs gènes et sont donc souvent létales dès le stade embryonnaire. Il a néanmoins été récemment décrit des variations à l'origine d'altérations de facteurs *trans* de l'épissage associées à des maladies spécifiques de certains tissus (Daguenet *et al.*, 2015 ; Scotti and Swansson, 2016).

Spinal muscular atrophy (SMA)	<i>SMN1</i> (REFS 136,137) (c.922+6 T/G), deletion	Loss of SMN full-length protein	Altered RNP biogenesis ⁹⁰	Recessive
Amyotrophic lateral sclerosis (ALS)	<i>TARDP</i> ⁷⁷ (c.991C>A), (c.1009A>G)	C-terminal mutations alter protein-protein interactions	TDP-43 target mis-splicing	Sporadic and Dominant
	<i>FUS</i> ¹³⁵ (c.1566C>T), (c.1561T>G)	• Decreased U1 interaction • Increased SMN binding	FUS target mis-splicing	Dominant
Dilated cardiomyopathy (DCM)	<i>RBM20</i> (REF. 139) (c.1962T>G)	Altered R/S RNA binding domain	<i>TTN</i> mis-splicing	Dominant
Limb-girdle muscular dystrophy 1G (LGMD1G)	<i>HNRPDL</i> ¹⁴⁰ (c.1667G>A), (c.1667G>C)	Altered import of HNRPDL into nucleus	HNRPDL target mis-splicing	Dominant
Autosomal dominant leukodystrophy (ADLD)	<i>LMNB1</i> (REF. 141) duplication	Increased <i>RAVER2</i> expression	PTBP1 target mis-splicing mediated by <i>RAVER2</i>	Dominant

Tableau 16. Exemples de variations associées à une altération d'un facteur *trans* d'épissage et responsable d'une maladie mendélienne (d'après Scotti and Swansson, 2016)

Les défauts d'épissage survenant à la suite d'une anomalie d'un composant du splicéosome peuvent être engendrés par des mutations touchant des gènes codant des facteurs essentiels du splicéosome ou codant des protéines responsables de la biogenèse de la machinerie d'épissage (Wang and Cooper, 2007 ; Daguenet *et al.*, 2015). L'altération d'un gène codant un composant essentiel du splicéosome a par exemple été observée dans le cas du syndrome de Nager, un syndrome rare de malformations congénitales (Xiong and Li, 2020). La première description de ce syndrome remonte à plus de 70 ans et, jusqu'à récemment, aucun gène de prédisposition n'avait été associé à cette maladie. C'est seulement en 2012 que, grâce à des efforts internationaux et à l'avènement des nouvelles technologies de séquençage à haut débit, l'analyse des premiers exomes de patients atteints de cette maladie a permis d'identifier des anomalies causales de type *frameshift* dans le gène *SF3B4* (Bernier *et al.*, 2012). Ce gène code la sous-unité 4 du complexe SF3b4, composant essentiel de la ribonucléoprotéine nucléaire snRNP U2 qui, en interagissant avec la protéine SF3b145, permet la

liaison de snRNP2 au point de branchement. Ainsi, les variations identifiées chez les patients sont responsables de la perte de fonction du complexe SF3b4 et bloquent la liaison de snRNP2 au point de branchement. La mise en cause de l'altération d'autres composants du spliceosome a été établie dans le cas de certaines maladies mendéliennes, par exemple, de désordres hématopoïétique, lymphoïdes, myéloïdes, de syndromes malformatifs multiples ou encore de syndromes dégénératifs (Daguenet *et al.*, 2015). Un exemple emblématique de l'altération de la biogenèse de la machinerie d'épissage est celui de l'amyotrophie spinale. Comme décrit plus haut, cette maladie résulte de la perte de fonction de la protéine SMN essentiellement codée par le gène *SMN1*. Cette protéine a pour fonction majeure de réguler la biogenèse et l'assemblage des snRNP. La perte de SMN bouscule donc l'organisation du spliceosome et ainsi le bon fonctionnement de l'épissage (Coady and Lorson, 2011). D'autres troubles neurodégénératifs sont également associés à des altérations de la biogenèse de la machinerie d'épissage notamment, la sclérose latérale amyotrophique également appelée maladie de Charcot. Cette maladie neurodégénérative très grave est associée à de nombreuses mutations localisées dans près d'une vingtaine de gènes différents parmi lesquels *TARDBP* et *FUS*. Les protéines codées par ces gènes, respectivement TDP-43 et FUS, sont capables d'interagir avec les ARN constituant les snRNPs afin de réguler le spliceosome (Tsuiji *et al.*, 2013 ; Barmada *et al.*, 2015). Une anomalie au sein d'un des gènes codant ces protéines conduit alors à une altération des taux des composants du spliceosome et de ce fait, à un défaut d'épissage dans transcrits cibles.

IV.2. Les outils de prédiction de l'impact des variations splicéogéniques

Depuis l'avènement des nouvelles technologies de séquençage, le nombre de variations identifiées au titre du diagnostic ne cessent d'augmenter, rendant impossible une analyse fonctionnelle pour chacune de ces variations, du fait du temps requis et du coût important de ce type d'analyse. A titre d'exemple, les bases de données de variations identifiées dans le génome des patients, telles que ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>, Landrum *et al.*, 2014, 2018) et HGMD (<http://www.hgmd.cf.ac.uk/ac/index.php>, Stenson *et al.*, 2017), comptabilisaient en décembre 2020 plus de 850 000 et 270 000 variations uniques, respectivement. Afin de sélectionner en priorité les variations les plus susceptibles d'impacter l'épissage pour des analyses expérimentales parmi la pléthore de variations identifiées, il est actuellement nécessaire de réaliser une première stratification des mutations au moyen d'outils bioinformatiques (Tableau 17). De nombreux algorithmes *in silico* ont été implémentés afin d'appréhender l'impact potentiel des variations sur les différents signaux *cis* sous-tendant le processus d'épissage (Spurdle *et al.*, 2008).

Outils	Motifs	Interface Web	Références
SSF	5'/3'ss	ND	Shapiro and Senapathy, 1987
NNSplice	5'/3'ss	https://www.fruitfly.org/seq_tools/splice.html	Reese <i>et al.</i> , 1997
MES	5'/3'ss	http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html	Yeo and Burge, 2004
HSF	5'/3'ss	https://www.genomnis.com/access-hsf	Desmet <i>et al.</i> , 2009
SPICE	5'/3'ss	https://sourceforge.net/projects/spicev2-1/	Leman <i>et al.</i> , 2018
SpliceAI	5'/3'ss	https://pypi.org/project/spliceai/	Jaganathan <i>et al.</i> , 2019
SVM-Bpfinder	BS	http://regulatorygenomics.upf.edu/Software/SVM_BP/	Corvelo <i>et al.</i> , 2010
BBP	BS	https://github.com/zhqingit/BPP	Zhang <i>et al.</i> , 2017
Branchpointer	BS	https://bioconductor.org/packages/release/bioc/html/branchpointer.html	Signal <i>et al.</i> , 2018
LaBranchoR	BS	http://bejerano.stanford.edu/labranchor/	Paggi and Bejerano, 2018
RNABPS	BS	http://nscibio.jbnu.ac.kr/tools/RNABP/	Nazari <i>et al.</i> , 2019
QUEPASA	ESR	http://bioinfo.univ-rouen.fr/HExoSplice_submit/	Ke <i>et al.</i> , 2011 ; Di Giacomo <i>et al.</i> , 2013
HEXplorer	ESR	https://www2.hhu.de/rna/html/hexplorer_score.php	Erkelenz <i>et al.</i> , 2014
SPANR	ESR	http://tools.genes.toronto.edu/	Xiong <i>et al.</i> , 2015
HAL	ESR	http://splicing.cs.washington.edu/	Rosenberg <i>et al.</i> , 2015
SPiP	5'/3'ss, BS, ESR	à venir	à venir

Tableau 17. Outils bioinformatiques de prédiction de l'impact des variations splicéogéniques (adapté de Canson *et al.*, 2020)

IV.2.1. Outils de prédiction dédiés aux sites d'épissage

Dans la mesure où les séquences consensus des sites 5' et 3' d'épissage sont aujourd'hui relativement bien caractérisées, il n'est pas étonnant de constater que de nombreux outils ont été développés dans le but de déterminer la position et la force de ces signaux dans une séquence donnée (Spurdle *et al.*, 2008 ; Canson *et al.*, 2020). Parmi les outils les plus largement utilisés, on peut citer SSF-Like (SSF-L, développé par Interactive Biosoftware à partir de *Splice Site Finder*, SSF, Shapiro and Senapathy, 1987), NNSplice (*Neural Network Splice*, Reese *et al.*, 1997), MES (*MaxEntScan*, Yeo and Burge, 2004) et HSF (*Human Splicing Finder*, Desmet *et al.*, 2009) (Tableau 17). Ces différents outils reposent tous sur l'utilisation des informations acquises sur les séquences consensus des sites 5' et 3' d'épissage et pourtant, les scores de prédiction générés par chaque outil divergent. En effet, chaque outil est associé à une définition des sites d'épissage quelque peu différente qui se traduit par l'application d'un algorithme qui lui est propre (Shapiro and Senapathy, 1987 ; Reese *et al.*, 1997 ; Yeo and Burge, 2004 ; Desmet *et al.*, 2009). Ainsi, les scores établis par les outils SSF et HSF sont générés par une matrice de position pondérée (PWM, *Position Weight Matrix*) obtenue par alignement de séquences correspondant à un jeu de données de jonctions exons/introns constitutives. Chaque nucléotide des séquences consensus retrouvées se verra attribuer une fréquence par rapport au jeu de données et la somme de ces fréquences permettra d'établir un score de la force du site d'épissage identifié dans une séquence donnée (Shapiro and Senapathy, 1987 ; Desmet *et al.*, 2009). La version modifiée de SSF, SSF-L permet le calcul des scores des 5'ss à la fois pour les motifs canoniques GT mais

également pour les motifs GC. Une des principales limites de cette analyse est le fait que chaque nucléotide composant le motif donné est traité indépendamment des autres. Afin de prendre en compte l'interaction potentielle des nucléotides entre eux, l'outil MES a été développé. Cet algorithme repose sur la méthode d'entropie maximale qui, en utilisant un large jeu de données de sites d'épissages correspondant à des séquences consensus, établit la somme des probabilités des séquences possibles (Yeo and Burge, 2004). Ce calcul est défini par un ensemble de contraintes dont notamment la position des nucléotides et l'interdépendance des nucléotides de proche en proche ou bien l'interaction des nucléotides adjacents à un nucléotide d'écart ou plus. Enfin, les scores établis par l'outil NNSplice reposent sur un autre type d'analyse. Il s'agit d'une méthode statistique appelée *neural network* en référence au réseau neuronal du cerveau humain où chaque « neurone » est assimilé à une fonction mathématique (Reese *et al.*, 1997). Dans ce type d'analyse les « neurones » sont organisés en couches où chacune d'entre elles intègre les données de la couche précédente et ce afin d'aboutir à un calcul final générant le score de prédiction de la force du site 5' ou 3' d'épissage.

Quel que soit l'algorithme appliqué pour la détermination des scores, tous les outils fonctionnent sur le même principe : plus la valeur du score obtenu est grande, plus le site d'épissage est associé à une probabilité élevée d'être reconnu par le splicéosome. Néanmoins, les fenêtres ou échelles de scores sont différentes d'un outil à l'autre. Ainsi, pour SSF-L et HSF, les scores varient de 0 à 100 tandis que NNSplice attribue un score allant de 0 à 1 et MES, de 0 à 16 selon l'interface Alamut Visual. Ces outils permettent non seulement d'appréhender la force de chaque site d'épissage et ainsi d'évaluer la probabilité que ce site soit effectivement reconnu par le splicéosome, mais également de comparer la force d'un site d'épissage dans le contexte sauvage ou muté (Hartmann *et al.*, 2008). Si la force prédite du site d'épissage est fortement diminuée voire nulle du fait de la variation, il est alors estimé que la variation a une forte probabilité d'affecter l'épissage. De façon intéressante, ces outils prédisent non seulement la destruction ou la diminution, de la force d'un site d'épissage constitutif, mais ils sont également capables de prédire la création ou l'activation de sites cryptiques à distance des jonctions naturelles exons/introns (Houdayer *et al.*, 2012). Dans ce cas, l'éventuelle utilisation d'un tel site par la machinerie d'épissage peut potentiellement être appréciée par comparaison de la force de ce site avec celle du site constitutif.

Ces outils bioinformatiques ont fait l'objet de plusieurs études visant à vérifier la concordance des prédictions émises avec la réalité des données expérimentales. De nombreuses variations ont ainsi été analysées, notamment dans les gènes *RB1*, *CFTR*, *MMR* et *BRCA* (Sharp *et al.*, 2004 ; Sharma *et al.*, 2014 ; Tournier *et al.*, 2008 ; Théry *et al.*, 2011 ; Houdayer *et al.*, 2012). Il est alors apparu que ces outils présentaient tous une valeur prédictive importante, en accord avec les données expérimentales. Par

conséquent, les laboratoires de recherche et de diagnostic moléculaire s'appuient désormais sur cette palette d'outils capable de prédire l'effet potentiel des variations sur les éléments *cis* d'épissage et ce, afin de sélectionner en priorité les variations qui feront l'objet d'une analyse expérimentale (Spurdle *et al.*, 2008 ; Hartmann *et al.*, 2008). Ceci est d'autant plus facilité qu'une partie de ces outils peuvent être interrogés indépendamment directement sur leur site en libre accès (sauf SSF ou SSF-L) ou simultanément, notamment à l'aide de l'interface graphique Alamut Visual développée à Rouen par la société Interactive Biosoftware. Bien que ces quatre outils soient tous associés à une bonne concordance de leur prédiction avec les données expérimentales, il est apparu important de les comparer entre eux afin d'identifier les approches bioinformatiques présentant la meilleure capacité prédictive. Ainsi, une étude menée par le groupe GGC-épissage en 2012 a comparé les scores prédictifs obtenus par chaque outil avec les résultats expérimentaux de l'effet sur l'épissage de 272 variations localisées dans les gènes *BRCA1* et *BRCA2*, obtenus rétrospectivement par les laboratoires du réseau national d'oncogénétique (Houdayer *et al.*, 2012). Ces travaux ont permis de montrer que les outils de prédiction MES puis SSF-L sont associés à la meilleure capacité prédictive. Il est d'ailleurs à noter que depuis cette étude, l'outil MES a fait l'objet d'analyses supplémentaires confirmant son excellente capacité de prédiction (Sharma *et al.*, 2014 ; Leman *et al.*, 2018 ; Montalban *et al.*, 2019). L'étude de 2012 a non seulement permis de démontrer quels outils de prédiction étaient les plus fiables, mais elle a également permis d'établir des seuils d'utilisation de ces outils. En effet, jusqu'en 2012, en l'absence de recommandations spécifiques, les seuils à partir desquels tout changement de scores émis par un de ces outils *in silico* était considéré comme prédictif d'un défaut d'épissage étaient établis de façon totalement arbitraire. Grâce à l'étude d'Houdayer et ses collaborateurs, des seuils correspondant à la valeur de la différence entre les scores observés dans les contextes sauvage et muté et avec lesquels une sensibilité et une spécificité optimales ont été obtenues (sensibilité = 96% et 91%, spécificité = 74% et 87%, pour MES et SSF, respectivement), ont ainsi pu être définis. Pour la première fois, des recommandations pour l'utilisation de ces prédictions bioinformatiques ont été émises. Elles proposent que les variations entraînant simultanément une diminution d'au moins 15% du score MES et de 5% du score SSF-L, sont fortement suspectées d'impacter l'épissage et devront faire l'objet d'une analyse expérimentale. La combinaison de ces 2 approches *in silico* permettant alors d'obtenir une sensibilité de 96% et une spécificité de 83% (Houdayer *et al.*, 2012).

Une nouvelle stratégie ayant pour objectif de rendre encore plus performantes les prédictions de l'impact potentiel des variations sur les sites d'épissage a été récemment développée en se basant sur une régression logistique représentant la meilleure combinaison possible des scores des deux outils *in silico* les plus fiables, MES et SSF-L (Leman *et al.*, 2018). Ce nouvel outil de prédiction a été validé à partir d'un jeu important de données collectées grâce à un travail collaboratif au sein du réseau

national GGC-épissage et avec le consortium international ENIGMA. C'est ainsi que les données expérimentales d'épissage de 395 variations (305 localisées dans les gènes *BRCA1* et *BRCA2* et 90 réparties dans les gènes *CFTR*, *CTRC*, *HFE*, *HJV*, *LRP5*, *PKD1*, *RHD*, *SLC40A1* et *TFR2*) ont été confrontées avec les prédictions *in silico*. Cette approche bioinformatique, nommée SPiCE (*Splicing Predictions in Consensus Elements*), a montré correctement prédit près de 99% des défauts d'épissage dans le jeu de validation. Sur la base de ces performances, ce nouvel outil est désormais recommandé pour une utilisation dans un contexte de diagnostic pour la prédiction de l'impact éventuel des variations localisées dans les régions consensus des sites d'épissage définies par cet outil (-3 à +8 pour le 5'ss ; -12 à +2 pour le 3'ss). Il est important de noter que par définition l'outil MES n'est pas en capacité de prédire l'utilisation des 5'ss de type GC. En effet, les prédictions émises par cet outil pour les substitutions T>C localisées en position +2 (GT>GC) indiquent une destruction totale du 5'ss alors que celui-ci est utilisé dans 15 à 18% des cas conduisant alors à des taux de transcrits normaux de 1 à 89% (Lin *et al.*, 2019).

Plus récemment, l'implémentation de stratégies d'apprentissage profond ou *deep learning*, une méthode mathématique reprenant le principe du *neural network* a également permis le développement d'un nouvel outil appelé SpliceAI basé sur l'intelligence artificielle (Jaganathan *et al.*, 2019). Contrairement aux outils bioinformatiques précédemment cités qui prennent en compte uniquement les courtes séquences consensus autour de la variation, SpliceAI analyse, pour chaque position dans le pré-ARNm, les 10 000 nucléotides flanquants afin de prédire si cette position correspond à un site d'épissage potentiel. La force de cet outil repose également sur sa bonne capacité à prédire les sites cryptiques susceptibles d'être utilisés en relai du site naturel. La bonne performance de SpliceAI a récemment été confirmée par plusieurs travaux indépendants (Wai *et al.*, 2020 ; Chen *et al.*, 2020 ; Saint-Martin *et al.*, 2021).

IV.2.2. Outils de prédiction dédiés au site de branchement

Contrairement aux sites 5' et 3' d'épissage aujourd'hui bien définis, les séquences consensus des sites de branchement restent encore mal connues et difficiles à analyser étant donné la dégénérescence élevée de la région autour du point de branchement et la petite taille de ce motif consensus (Gao *et al.*, 2008). De plus, les sites de branchement sont localisés à une distance variable (le plus souvent 18 à 44 nucléotides) en amont des sites d'épissage, rendant ainsi leur prédiction particulièrement ardue. Néanmoins, la majorité des points de branchement sont caractérisés par la séquence consensus YUnAY (le A indiquant le point de branchement proprement dit) et c'est sur la base de ce motif qu'ont été développés les premiers outils *in silico* de prédiction de la position, voir de

l'altération potentielle, de ce site. On peut notamment citer SROOGLE (Schwartz *et al.*, 2009) et HSF (*Human Splicing Finder*, Desmet *et al.*, 2009). Toutefois, une limite majeure de ces outils est le fait que leur fiabilité n'a pu être évaluée de façon exhaustive étant donné le faible nombre de variations connues comme étant à l'origine d'une altération du BP (Lewandowska *et al.*, 2013 ; Bitton *et al.*, 2014).

Récemment, grâce au développement des techniques de séquençage à haut-débit, cette limite tend à être dépassée. En effet, depuis 2015, des études à large échelle ont été menées dans le but de caractériser finement les points de branchement du génome humain, via des approches de cartographie (*mapping*), et ont même permis d'identifier un plus grand nombre de variations à l'origine d'une altération du site de branchement (Mercer *et al.*, 2015 ; Chiang *et al.*, 2017 ; Zhang *et al.*, 2017 ; Signal *et al.*, 2018 ; Pineda and Bradley, 2018 ; Briese *et al.*, 2019). Les données issues de ces études ont été intégrées au sein d'algorithmes d'apprentissage automatique (*machine learning*) et ce afin de générer à la fois des outils d'annotation des points de branchement mais également des outils de prédiction de l'impact des variations sur le BP. On peut notamment citer SVM-BPfinder (*Support Vector Machine learning for Branch Points*, Corvelo *et al.*, 2010), BBP (*Branch Point Predictions*, Zhang *et al.*, 2017), Branchpointer (Signal *et al.*, 2018), LaBranchoR (*Long short-term memory network Branchpoint Retriever*, Paggi and Bejerano, 2018) et RNABPS (*RNA Branch Point Selection*, Nazari *et al.*, 2019) (Tableau 17). Les performances de ces outils différents ainsi que de celle de HSF ont très récemment été évaluées au cours d'une étude française en collaboration avec le consortium international ENIGMA (Leman *et al.*, 2020). L'utilisation d'un 3'ss étant conditionnée par la présence d'un point de branchement, Leman et ses collaborateurs ont mis à profit deux jeux de données : un premier répertoriant la totalité des 3'ss constitutifs humains décrits dans la base de données Ensembl (n= 264 787) et un second rassemblant une série de 3'ss alternatifs détectés par la technique de RNA-seq (n = 51 986). L'outil *in silico* Branchpointer s'est révélé être l'approche la plus efficace pour détecter les vrais points de branchement avec une exactitude supérieure à 99% (Leman *et al.*, 2020). Dans le but d'évaluer la capacité prédictive des différents outils, cette étude a également comparé les scores obtenus par chaque outil *in silico* avec les données expérimentales de l'impact sur l'épissage de 120 variations localisées dans la région du point de branchement (*i.e.* 18 à 44 nucléotides en amont des 3'ss). Ici, c'est l'outil de prédiction BPP qui s'est avéré le plus performant avec une exactitude de plus de 89% (Leman *et al.*, 2020).

IV.2.3. Outils de prédiction dédiés aux éléments de régulation de l'épissage

Tout comme le point de branchement, les éléments de régulation de l'épissage ne sont pas encore totalement bien caractérisés du fait de leur multiplicité et de leur mode d'action combinatoire et chevauchant. De plus, le rôle des SRE est dépendant de plusieurs paramètres parmi lesquels leur position exonique ou intronique, leur distance par rapport aux sites d'épissage, leur interdépendance ou encore l'expression tissu spécifique de certains facteurs de régulation de l'épissage, représentant ainsi une difficulté supplémentaire dans l'élaboration d'outils *in silico* de prédiction de ces éléments (Fairbrother *et al.*, 2004 ; Spurdle *et al.*, 2008 ; Ke and Chasin, 2011). Néanmoins, plusieurs outils bioinformatiques ont été développés dans le but de prédire la présence d'éléments de régulation de l'épissage au sein d'une séquence donnée tels que RESCUE-ESE (Fairbrother *et al.*, 2002), ESE-finder (Cartegni *et al.*, 2003) et EX-SKIP (Raponi *et al.*, 2011). Ces différentes approches reposent sur des analyses de l'enrichissement des motifs proches des sites d'épissage, sur des données de conservation entre les espèces ou encore sur des données expérimentales (Castle *et al.*, 2008 ; Goren *et al.*, 2006 ; Smith *et al.*, 2006). Toutefois, il a été démontré à partir de plusieurs études comparatives que les prédictions de l'impact des variations sur les éléments de régulation de l'épissage obtenues à l'aide de ces outils sont rarement concordantes avec les données expérimentales (Raponi *et al.*, 2007 ; Arnold *et al.*, 2009 ; Soukariéh *et al.*, 2016). Sur la base de ces travaux, il apparaît donc impossible d'utiliser ces outils dans le but de stratifier les variations pour les tests fonctionnels d'épissage dans un contexte de diagnostic. Cependant, de nouvelles approches ont récemment été développées (Tableau 17).

Une première étude a évalué expérimentalement l'activité potentielle en tant qu'ESR de toutes les combinaisons possibles d'hexamères dans un système minigène (Figure 30 ; Ke *et al.*, 2011). Cette analyse a ainsi permis d'attribuer un score ESRseq à chacune des 4096 séquences hexamériques possibles. Un score ESRseq positif est alors prédictif d'un élément ESE favorisant l'inclusion de l'exon, un score nul d'un élément neutre et enfin, un score négatif d'un élément ESS susceptible de contribuer à l'exclusion de l'exon. Notre groupe a ensuite dérivé ces résultats afin d'appréhender l'impact potentiel d'une variation exonique sur un ESR en calculant la différence entre la somme des scores ESRseq des 6 hexamères chevauchant la position nucléotidique d'intérêt dans le contexte muté versus le contexte sauvage ($\Delta tESR_{seq}$, *total ESRseq score change* = $tESR_{seq}(mut) - tESR_{seq}(WT)$) (Figure 30 ; Di Giacomo *et al.*, 2013).

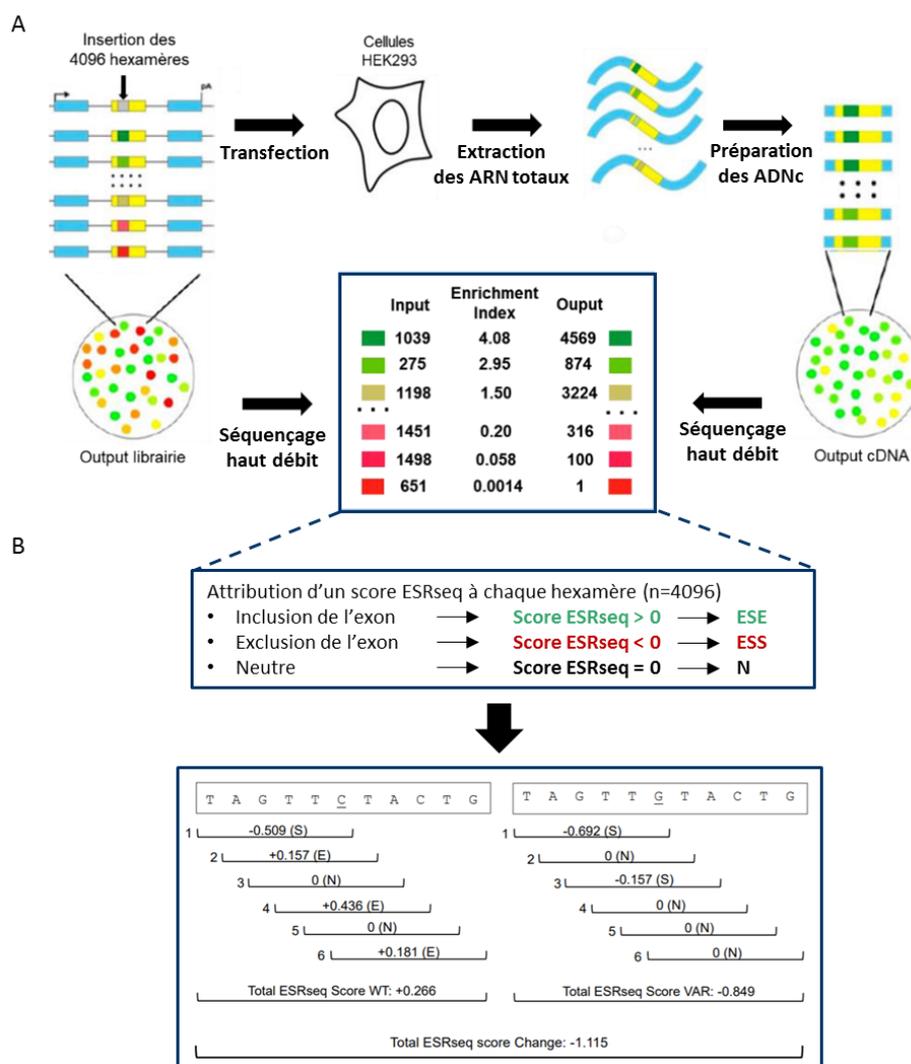


Figure 30. Méthode de prédiction de l'altération des éléments régulateurs de l'épissage basé sur le calcul du score $\Delta tESR_{seq}$ (adapté de Ke *et al.*, 2011 ; Di Giacomo *et al.*, 2013). A. Evaluation expérimentale de toutes les combinaisons possibles d'hexamères et attribution d'un score ESRseq pour chacun d'entre eux. B. Calcul du score $\Delta tESR_{seq}$ au niveau d'une position exonique donnée. L'exemple présenté ici est celui de la variation c.617C>G dans l'exon 7 de *BRCA2*.

Une seconde approche, HEXplorer, a été développée peu après et repose sur les scores Z_{EI} qui dérivent d'une analyse statistique basée sur la distribution relative des hexamères dans les exons et les introns en fonction de leurs distances aux sites d'épissage (Erkelenz *et al.*, 2014). Ainsi, comme pour le calcul du score $\Delta tESR_{seq}$, le calcul du score ΔHZ_{EI} correspond à la différence entre l'allèle muté et l'allèle sauvage du total des scores HZ_{EI} des 6 hexamères chevauchant la position de la variation d'intérêt. Un an après la mise au point de cette méthode statistique, une troisième approche a été développée en se basant quant à elle sur une méthode de *machine learning*. L'approche SPANR (*splicing-based analysis of variants*) repose sur la compilation de données obtenues par RNA-seq à partir de 16 tissus humains sains différents (Xiong *et al.*, 2015). L'ensemble de ces données a permis

d'entraîner un modèle informatique d'épissage capable de prédire le taux de transcrits ayant inclus l'exon (Ψ) dans un contexte sauvage ou muté. Par comparaison entre les contextes sauvage et muté ($\Delta\Psi$ -SPANR), cette approche permet ainsi d'évaluer l'effet de toutes variations exoniques et introniques localisées au niveau des éléments de régulation mais également au niveau des sites d'épissage. Enfin au cours de la même année, une quatrième approche, HAL (*Hexamer Additive Linear*), a été développée en se basant également sur une méthode de *machine learning* mais aussi sur une technique de biologie synthétique (Rosenberg *et al.*, 2015). L'équipe de Rosenberg et ses collaborateurs ont généré près de 2 millions de minigènes synthétiques et mesuré, par la technique de RNA-seq, les ratios des différentes isoformes résultant de l'épissage de ces minigènes. Au même titre que l'approche SPANR, l'approche HAL calcule le pourcentage de transcrits ayant inclus l'exon (Ψ), permettant ainsi de générer un score prédictif ($\Delta\Psi$ -HAL) de l'impact d'une variation sur les ESR.

Afin d'évaluer le pouvoir prédictif de ces différents outils de prédiction dédiés principalement aux ESR, une première étude a été réalisée en 2016 par notre groupe sur trois de ces quatre nouvelles approches : QUEPASA (Δt ESRseq), HEXplorer (Δ HZEI) et SPANR ($\Delta\Psi$ -SPANR) (Soukarieh *et al.*, 2016). Sur la base de 5 jeux de données rassemblant les effets sur l'épissage de 154 variations à distance des sites d'épissage dans certains exons choisis comme modèles (*MLH1* exon 10, *BRCA2* exon 7, *BRCA1* exon 6, *CFTR* exon 12 et *NF1* exon 37), il a été montré une très bonne concordance entre les prédictions issues de ces nouveaux outils et les résultats expérimentaux, notamment pour les scores Δt ESRseq et Δ HZEI. Il est intéressant de noter que les scores Δt ESRseq et Δ HZEI sont non seulement capables de prédire si une variation exonique est susceptible d'altérer l'épissage via une modification des ESR, mais elles sont également en capacité de corrélérer le niveau d'inclusion de l'exon avec les valeurs des scores prédictifs. En vue d'inclure ces approches dans les stratégies de stratification des variations exoniques à analyser en priorité dans les tests fonctionnels d'épissage, il était nécessaire d'évaluer ces outils à plus large échelle. C'est dans ce but qu'une récente étude de notre groupe a évalué la capacité prédictive des quatre approches citées ci-dessus avec les données expérimentales extraites de la littérature de plus de 1300 variations localisées dans 89 gènes de prédisposition à des maladies mendéliennes (Tubeuf *et al.*, 2020b). Ces travaux ont permis de confirmer la fiabilité des approches QUEPASA et HEXplorer mais ils ont également révélé que les approches SPANR et HAL sont elles aussi associées à une bonne capacité prédictive. De plus, les analyses statistiques ont démontré que ces quatre outils sont capables de prédire les variations les plus susceptibles d'impacter les ESR, mais également la direction des défauts d'épissage induits, ainsi que la sévérité de ces anomalies. Tubeuf et ses collaborateurs ont ainsi établi des seuils de décision optimaux pour chaque outil afin d'optimiser la performance des analyses bioinformatiques dédiées aux ESR. Il est d'ailleurs apparu que l'approche QUEPASA présente la plus haute sensibilité dans la prédiction des variations conduisant à un saut

d'exon tandis que l'approche HAL est associée à la sensibilité la plus élevée quand il s'agit de prédire l'augmentation de l'inclusion de l'exon. Enfin, il est apparu que la combinaison de QUEPASA et de HAL permet d'améliorer la spécificité de l'analyse.

Ces nouvelles approches bioinformatiques semblent donc très prometteuses pour la détection des variations à l'origine d'une altération d'un élément de régulation de l'épissage. Il apparaît donc essentiel de poursuivre l'évaluation de ces outils sur de plus larges jeux de données. Il s'agit de définir notamment quels sont les critères susceptibles d'affecter la fiabilité de ces prédictions d'ESR en prenant en considération notamment la taille de l'exon et des introns flanquants, la force des sites d'épissage ainsi que la densité en ESR prédits. Une étude récente a démontré l'intérêt des données issues d'analyses de saturation du génome (SGE, *Saturation Ggenome Editing*) (Findlay *et al.*, 2018) dans la cartographie d'ESRs potentiels du gène *BRCA1* (Canson *et al.*, 2020). Une cartographie précise du paysage des éléments de régulation dans un gène donné représente aujourd'hui un challenge important.

IV.2.4. Outils de prédiction combinés

La plupart des outils bioinformatiques précédemment cités sont dédiés à la prédiction de l'impact des variations sur un type de signaux d'épissage spécifiques. L'analyse *in silico* de l'impact potentiel d'une variation sur l'épissage requiert donc, en théorie, d'utiliser de façon séquentielle chacun des différents outils pour appréhender l'effet de la variation sur les multiples signaux d'épissage. De plus, ces différentes analyses sont potentiellement nécessaires pour anticiper un effet combiné d'une variation, par exemple du fait de son action conjointe sur les sites d'épissage, d'une part, et les ESR, d'autre part. C'est la raison pour laquelle Leman (membre de notre groupe) et ses collaborateurs ont développé un outil nommé SPiP (*Splicing Prediction Pipeline*) (Leman *et al.*, manuscrit en cours de préparation). Cet outil *in silico* repose sur un algorithme décisionnel intégrant les outils optimaux pour chaque signal d'épissage. Ainsi, c'est l'outil de prédiction SPiCE qui a été sélectionné pour prédire les altérations des sites d'épissage, l'outil BPP pour prédire les modifications du point de branchement, l'outil MES pour prédire les anomalies de la région riche en pyrimidine et enfin, l'approche QUEPASA pour prédire les altérations des éléments de régulation de l'épissage. L'outil SPiP a été validé à partir d'un jeu de données de plus de 3 000 variations localisées dans 218 gènes. De façon tout à fait remarquable, cette approche combinée permet de prédire les différents types d'altérations de l'épissage (saut d'exon, utilisation d'un nouveau site d'épissage ou d'un site d'épissage cryptique, pseudo-exonisation) et est associée à une précision de plus de 90% ainsi qu'une sensibilité et une spécificité toutes deux également supérieures à 90%. Cet outil de décision à large

échelle se révèle donc particulièrement puissant dans la stratification des variations pour les tests fonctionnels d'épissage.

Malgré l'évolution permanente des outils *in silico* de prédiction et l'amélioration continue de leurs performances, ces approches n'en restent pas moins limitées car incapables d'identifier précisément les conséquences des anomalies d'épissage au niveau des transcrits d'un point de vue qualitatif et quantitatif. En effet, les prédictions générées à l'aide de ces outils ne permettent généralement pas de définir avec certitude, d'une part, la nature exacte du ou des transcrits aberrants produits dans le contexte muté et, d'autre part, le degré de sévérité ou l'amplitude des défauts d'épissage. Ces informations essentielles pour l'interprétation biologique et clinique des variations ne peuvent être acquises que par le biais d'une analyse expérimentale. Du fait du débit relativement faible de ces approches expérimentales, les outils bioinformatiques de prédiction sont aujourd'hui mis à profit pour établir une première stratification des variations à analyser dans les tests fonctionnels d'épissage. A l'avenir, le développement et l'optimisation d'outils combinés pourrait simplifier l'étape d'analyse bioinformatique des variations génétiques dans un contexte clinique, jusqu'ici alourdie par l'utilisation séquentielle d'une multitude d'outils.

IV.3. Les approches expérimentales de détection des anomalies d'épissage

Les approches expérimentales de détection des anomalies d'épissage induites par des variations génétiques reposent sur une analyse comparative des profils d'épissage des transcrits exprimés dans le contexte sauvage avec ceux produits dans le contexte muté. Différents essais ont ainsi été développés avec pour objectif de caractériser l'effet des variations sur l'épissage en identifiant la nature des transcrits engendrés, d'une part, et en déterminant le degré de sévérité de l'altération, d'autre part. Ces données de nature qualitative et semi-quantitative sont essentielles pour appréhender les conséquences fonctionnelles et donc le caractère pathogène potentiel des mutations splicéogéniques. Il est ici important de distinguer les analyses expérimentales qui sont conduites à partir du matériel biologique du patient des méthodes alternatives qui permettent de contourner la disponibilité de ce type d'échantillons. Ces deux approches présentent chacune des avantages et des limites spécifiques et sont de ce fait souvent complémentaires.

IV.3.1. Analyses expérimentales à partir du matériel biologique des patients

Idéalement, l'analyse de l'effet d'une variation sur l'épissage doit se faire à partir de l'ARN du tissu impacté afin d'étudier l'expression endogène naturelle des transcrits d'intérêt dans le contexte tissulaire le plus pertinent vis-à-vis de la pathologie en question. En effet, l'épissage constitutif et

alternatif de l'ARN étant un mécanisme susceptible d'être régulé différemment selon le tissu considéré, une même variation peut être associée à des effets différents sur l'épissage selon le tissu analysé (Xu *et al.*, 2002 ; Baralle and Baralle, 2005). Toutefois, les tissus d'intérêt (sains et/ou tumoraux), notamment les tissus mammaires et ovariens dans le contexte du syndrome sein-ovaire et le tissu colorectal et de l'endomètre dans le contexte du syndrome de Lynch, sont rarement disponibles. En outre, même si un échantillon d'ARN du tissu d'intérêt est disponible, les transcrits aberrants ont potentiellement fait l'objet d'une dégradation par le système NMD rendant ainsi difficile la détection de défauts d'épissage à l'origine de PTCs. Le tissu tumoral peut représenter une alternative intéressante car plus facile d'accès que le tissu sain d'intérêt. Néanmoins, selon les conditions de prélèvement et de conservation de la tumeur et le type de fixation, la qualité de l'ARN peut être altérée, rendant ainsi difficile une quelconque analyse du profil d'épissage des transcrits. Il est également important de souligner que le tissu tumoral acquiert de nombreuses modifications génétiques somatiques au cours de la tumorigenèse (Graham and Sottoriva, 2017). Plusieurs études ont montré que le processus d'épissage est modifié dans les tumeurs et cette altération de l'épissage est susceptible de contribuer à la tumorigenèse (Escobar-Hoyos *et al.*, 2019 ; Rahman *et al.*, 2020). Dans le cadre d'une analyse de l'effet d'une variation constitutionnelle sur l'épissage à partir du tissu tumoral, il est donc potentiellement difficile de relier directement, de façon causale, les altérations observées à la mutation.

Du fait de l'absence récurrente d'échantillons de tissu relevant et/ou des inconvénients liés à une analyse à partir de la tumeur, l'étude des profils d'épissage à partir de l'ARN du patient est le plus souvent effectuée, par les laboratoires de diagnostic, à partir d'un type tissulaire différent de celui impacté, lorsque le gène étudié s'exprime dans ce matériel (Baralle *et al.*, 2009). C'est, par exemple, le cas de l'étude de l'impact des variations sur le gène *CFTR* qui repose aujourd'hui sur une analyse de l'ARN extrait à partir de cellules épithéliales de la muqueuse sinusienne. En effet, ce tissu présente une activité protéique de *CFTR* similaire à celle observée dans les tissus bronchiques, tissus touchés chez les patients atteints de mucoviscidose, et représente donc une source d'échantillons biologiques moins invasive et plus simple à obtenir que les tissus bronchiques (Brewington *et al.*, 2018). Le sang périphérique des patients constitue également une alternative intéressante à condition que le gène d'intérêt y soit exprimé, ce qui n'est notamment pas le cas du gène *CFTR*. A l'inverse, dans un contexte de gènes présentant une expression ubiquitaire comme les gènes *BRCA* ou *MMR*, ce matériel biologique représente une très bonne option en vue de l'analyse de l'impact des variations sur le profil d'épissage des transcrits d'intérêt. En effet, la collecte de sang périphérique dans des tubes contenant une solution de stabilisation des ARNs (tubes PAXgene, QIAGEN) est simple, rapide et permet d'obtenir de l'ARN de bonne qualité. Toutefois, la quantité d'ARN extraite à partir de ce type de matériel est très

limitée, ce qui est susceptible de représenter une limite pour des analyses multiples. De plus, le niveau d'expression de certains gènes, comme *BRCA1/2*, est relativement faible dans ces échantillons, ce qui implique d'utiliser une quantité importante d'ARN et d'adapter la technique d'analyse pour la rendre plus sensible (Baralle *et al.*, 2009 ; Baralle and Buratti, 2017). En outre, les transcrits aberrants porteurs d'un PTC peuvent faire l'objet d'une dégradation par le système NMD, compliquant alors l'interprétation des résultats obtenus à partir de ce type de matériel. Afin de contourner ce problème, un autre type de matériel biologique peut être utilisé : les lignées lymphoblastoïdes établies à partir de leucocytes du sang de patients et rendues immortelles par infection avec le virus EBV (*Epstein-Barr Virus*). Les lignées lymphoblastoïdes étant cultivables, elles peuvent faire l'objet d'un traitement par des inhibiteurs de la synthèse protéique, tels que la puromycine ou la cycloheximide, qui indirectement vont conduire à l'inhibition du NMD, cette voie de dégradation étant couplée au premier cycle de la traduction (Brojna and Wen, 2009). Cette approche permet ainsi d'empêcher la dégradation des transcrits porteurs d'un PTC, rendant de ce fait leur détection possible (Baralle *et al.*, 2009 ; Baralle and Buratti, 2017). De surcroît, la nature cultivable de ce matériel permet de contourner le problème de la quantité très limitée d'ARN obtenu à partir de sang frais collecté dans des tubes PAXgene. Cependant, l'immortalisation d'une lignée lymphoblastoïde est particulièrement longue, coûteuse et peut parfois échouer. De plus, même si cela est peu probable, il n'est pas exclu que certaines altérations d'expression puissent être dues au processus d'immortalisation et/ou à un grand nombre de passage des cellules (Oh *et al.*, 2013 ; Schafer *et al.*, 2013). Ainsi, le choix de tel ou tel autre matériel biologique n'est pas toujours évident car chaque type de matériel est associé à des avantages et des inconvénients qui leur sont propres. Il est également important de noter que, quel que soit le matériel biologique issu du patient utilisé, l'analyse et l'interprétation des données seront toujours compliquées par le contexte bi-allélique. En effet, les variations étudiées sont généralement à l'état hétérozygote dans le génome des patients. De plus, le lien de causalité entre les défauts d'épissage observés et la variation n'est pas directement établi par cette approche.

IV.3.1.a. Analyse des profils d'épissage des transcrits par RT-PCR

Classiquement, l'étude de l'impact d'une variation sur les profils d'épissage des transcrits d'intérêt s'effectue grâce à une technique développée dans les années 1980, la RT-PCR (*Reverse-Transcriptase Polymerase Chain Reaction*). Cette approche repose sur l'analyse ciblée des ADN complémentaires synthétisés à partir des transcrits à l'aide d'amorces positionnées dans les exons en 5' (amorce sens) et 3' (amorce reverse) de l'exon cible potentiellement affecté par le défaut d'épissage induit par la variation (Hartmann *et al.*, 2008 ; Gaildrat *et al.*, 2012). Cette analyse est conduite en parallèle à partir du matériel biologique du même type obtenu à partir d'individus contrôles, non

porteurs de variations dans le gène analysé, et celui du patient. Après migration des produits de RT-PCR par électrophorèse sur gel d'agarose, les profils d'épissage obtenus pour ces différents échantillons sont comparés et les différentes bandes observées sont découpées et séquencées afin de déterminer la nature exacte des isoformes produites. De nombreuses variations splicéogéniques ont été identifiées à l'aide de cette technique à partir d'ARN de patients (Tournier *et al.*, 2008 ; Théry *et al.*, 2011 ; Gaildrat *et al.*, 2012 ; Di Giacomo *et al.*, 2013 ; Soukarieh *et al.*, 2016). Néanmoins, la simple migration des produits de RT-PCR sur gel d'agarose ne permet *a priori* ni de discriminer les transcrits présentant une faible différence de taille ni d'évaluer avec précision les taux relatifs des différents transcrits potentiellement produits notamment du fait de la formation possible d'hétéroduplexes entre les différents transcrits. Pour analyser pleinement l'impact d'une variation splicéogénique, il est essentiel d'évaluer la contribution de l'allèle muté dans la production respective des transcrits aberrants, d'une part, et des transcrits normalement épissés, d'autre part. Toutefois, le séquençage des différents produits de RT-PCR ne peut apporter qu'une réponse partielle et non quantitative à cette question. Afin de contourner ce problème, il est fortement recommandé depuis quelques années de compléter l'analyse par une électrophorèse sur capillaire en conditions dénaturantes des amplicons issus d'une RT-PCR fluorescente. L'interprétation des résultats obtenus requiert une connaissance approfondie du profil d'épissage constitutif du gène d'intérêt, notamment au regard des potentiels épissages alternatifs existants (Baralle *et al.*, 2009 ; Colombo *et al.*, 2014 ; Thompson *et al.*, 2015 ; Fackenthal *et al.*, 2016 ; Davy *et al.*, 2017).

Aujourd'hui, la RT-PCR classique est devenue une technique de routine dans les laboratoires de recherche et de diagnostic moléculaire car elle présente l'avantage d'être facilement réalisable. Ce n'est que récemment que des approches quantitatives et semi-quantitatives ont été intégrées à cette analyse afin d'étudier plus précisément l'impact des variations splicéogéniques. On peut notamment distinguer la RT-PCR *long-range* qui permet d'amplifier de très grands fragments (10 à 20 kb) voire le transcrit pleine longueur, grâce à l'utilisation d'une ADN polymérase adaptée (Morak *et al.*, 2019). Une autre application de la RT-PCR est celle dérivée de la PCR quantitative (qPCR) également appelée PCR en temps réel (*real-time* PCR), qui, par introduction d'un fluorochrome à chaque production d'amplicon, permet de suivre en continu l'amplification de l'ADNc (de la Hoya *et al.*, 2016). Ainsi, l'analyse de la cinétique complète de la réaction d'amplification permet d'obtenir une quantification précise des transcrits. Plus récemment, la PCR digitale (*droplet digital* PCR) a également été mise à profit pour quantifier encore plus précisément les transcrits présents dans un échantillon et se révèle être plus sensible que la qRT-PCR (Colombo *et al.*, 2018).

IV.3.1.b. Analyse de l'expression allélique

Dans le but d'appréhender la contribution précise de l'allèle muté dans la production respective des transcrits normaux et aberrants, des analyses d'expression allélique (ASE, *Allele Specific Expression*) à partir de l'ARN de patients ont été développées. L'une des méthodes les plus utilisées est celle basée sur l'extension d'amorces, autrement appelée SNUPE (*Single Nucleotide Primer Extension*) ou SNaPshot (Singer-Sam, 1994 ; Matyas *et al.*, 2002 ; Yan, 2005). Cette technique repose sur l'extension d'une amorce hybridée sur l'ADNc juste en 5' ou 3' de la variation exonique d'intérêt (amorce sens ou anti-sens, respectivement). L'extension de cette amorce via l'incorporation de ddNTPs fluorescents (didésoxyriboNucléotides TriPhosphates) complémentaires à la séquence d'intérêt permet de discriminer les transcrits produits à partir de l'allèle sauvage ou muté (Figure 31). Cette approche permet ainsi de détecter avec précision un éventuel déséquilibre allélique résultant de la dégradation par le NMD de l'allèle porteur d'un PTC mais également d'évaluer la sévérité du défaut d'épissage induit par la variation. Ces deux points ont d'ailleurs pu être mis en évidence dans le cas de variations localisées dans les gènes *BRCA* (Caux-Moncoutier *et al.*, 2009) et les gènes *MLH1* et *MSH2* (Bujakolva *et al.*, 2008 ; Soukarieh *et al.*, 2016). La stratégie expérimentale d'analyse de l'expression allélique par extension d'amorces n'est possible qu'en présence d'une variation exonique hétérozygote. L'éventuel déséquilibre allélique induit par une variation intronique ne peut donc être directement évalué. Toutefois, il est possible d'appréhender le déséquilibre allélique provoquée par une variation intronique en interrogeant un polymorphisme exonique hétérozygote (SNP, *Single Nucleotide Polymorphism*) à proximité de de la variation d'intérêt (Caux-Moncoutier *et al.*, 2009).

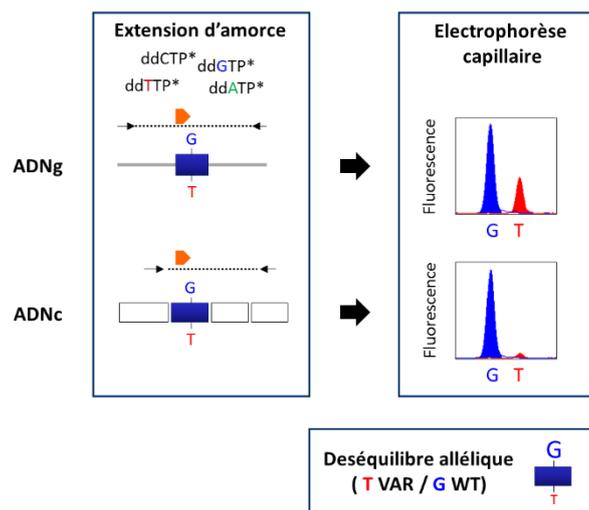


Figure 31. Principe de l'analyse de l'expression allélique par extension d'amorce. Le ratio allélique obtenu à partir des ADN complémentaires (ADNc) (produits de RT-PCR) est normalisé avec celui obtenu à partir des échantillons d'ADN génomique (ADNg) (produits de PCR) pour un même patient afin de s'affranchir du biais de fluorescence. L'exemple présenté ici illustre un déséquilibre allélique du fait de la perte d'expression de l'allèle muté (VAR) par rapport à l'allèle sauvage (WT).

L'analyse de l'expression allélique peut également être appréhendée par la méthode quantitative de pyroséquençage (Ahmadian *et al.*, 2005 ; Wang and Elbein, 2007 ; Gaildrat *et al.*, 2012 ; Yang *et al.*, 2016). Cette approche repose sur la détection des nucléotides incorporés lors d'une réaction enzymatique de synthèse d'ADN via la détection par chimioluminescence du pyrophosphate libéré au cours de la réaction. Au même titre que pour la méthode d'extension d'amorces, le pyroséquençage peut directement cibler la variation exonique d'intérêt dans l'ADNc du patient ou une variation exonique hétérozygote (SNP ou pas), rapporteur d'une variation intronique. Cette technique permet, d'une part, de détecter la perte d'expression d'un allèle comme cela a pu être démontré dans le cas d'individus porteurs d'une épimutation constitutionnelle du gène *MLH1* (Kwok *et al.*, 2010) et, d'autre part, de quantifier l'impact de variations exoniques sur les profils d'épissage des transcrits comme démontré dans le cas de variations exoniques du gène *BRCA2* (Gaildrat *et al.*, 2012).

IV.3.1.c. Analyses des défauts d'épissage par la technique de RNA-seq

En 2008, l'implémentation du séquençage à haut-débit de l'ARNm (RNA-seq) a représenté une révolution technologique pour l'analyse du transcriptome (Marioni *et al.*, 2008 ; Wang *et al.*, 2009). Cette approche présente un double avantage car elle génère des données qui permettent de déterminer la séquence de l'ensemble des transcrits exprimés dans un échantillon donné et d'estimer leurs taux relatifs. Le RNA-seq ciblé (Targeted RNA-seq, aussi appelé RNA CaptureSeq) est une approche dérivée visant à concentrer l'analyse uniquement sur les transcrits de certains gènes afin d'enrichir la couverture et d'augmenter la profondeur de lecture (Mercer *et al.*, 2011). Cet enrichissement rend alors possible la détection d'un plus grand nombre d'isoformes d'épissage, même les moins abondantes, pour un panel défini de gènes. Cette approche a été mise à profit récemment dans le but de caractériser les épissages alternatifs constitutifs des gènes *BRCA* et *MMR* (voir partie III.4.4.) (Davy *et al.*, 2017 ; Brandao *et al.*, 2019). Cette description détaillée de façon qualitative et quantitative des épissages alternatifs naturels dans ces gènes de prédisposition aux cancers est essentielle à l'interprétation biologique et clinique des variations splicéogéniques. De façon importante, ces travaux ont également permis de mettre en évidence le potentiel de cette technique dans la détection et la caractérisation qualitative et quantitative des événements aberrants et alternatifs d'épissage à partir de lignées lymphoblastoïdes de patients porteurs de variations splicéogéniques dans les gènes *BRCA1/2* (Davy *et al.*, 2017 ; Brandao *et al.*, 2019). En outre, le RNA-seq ciblé en présentant une résolution au niveau nucléotidique, permet à la fois d'identifier d'éventuelles anomalies d'épissage résultant de variations génétiques non identifiées mais également, d'analyser l'expression allélique (Han *et al.*, 2015).

La mise à profit de l'ensemble des approches mentionnées ci-dessus est dépendante de la disponibilité de l'ARN extrait à partir d'un matériel biologique de patient. De plus, l'interprétation des données issues de l'analyse des profils d'épissage des transcrits de patients est susceptible d'être complexifiée du fait de l'expression bi-allélique (WT/muté) et de la dégradation éventuelle des transcrits porteurs d'un PTC par le système NMD. Afin de s'affranchir de ces limites, les laboratoires ont développé des approches expérimentales alternatives capables d'évaluer l'effet de l'introduction de variations sur l'épissage dans des systèmes plus ou moins artificiels.

IV.3.2. Analyses expérimentales alternatives

IV.3.2.a. Test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes

C'est grâce à l'extension des méthodes de clonage moléculaire dans la seconde moitié du 20^{ème} siècle (Cohen *et al.*, 1973) que l'utilisation du minigène apparaît comme une technologie attractive. Aujourd'hui, les minigènes constituent un outil largement mis à profit par les laboratoires de recherche et de diagnostic moléculaire afin de caractériser finement l'effet des variations sur l'épissage (Baralle *et al.*, 2003 ; Cooper, 2005 ; Kishore *et al.*, 2008 ; Gaildrat *et al.*, 2010). En effet, cette méthode reprend un concept très simple : celui de la transfection transitoire dans des cellules humaines (ou de singe) de vecteurs d'expression comportant un promoteur constitutif (e.g. CMV ou SV40) capable d'induire la transcription d'une séquence d'intérêt insérée en aval de ce promoteur, dans ce cas particulier un minigène. Les minigènes sont, dans ce contexte, des séquences comprenant au moins deux exons séparés par un intron, représentant ainsi une version simplifiée d'un gène multi-exon. Le plus souvent, un ou plusieurs autres exons (hétérologues ou pas) avec leurs séquences introniques flanquantes peuvent alors être insérés au sein de l'intron du minigène « mère » résultant alors en un minigène à trois exons ou plus (Figure 32). Il s'agit d'une stratégie permettant d'étudier l'épissage d'exons dits « cassette » (exons internes) sauvages (WT) ou mutés. Le segment génomique WT ou muté, correspondant à l'exon ciblé ainsi que ses régions introniques flanquantes, peut être amplifié à partir de l'ADN génomique (ADNg) d'un patient (hétérozygote pour la variation étudiée) puis inséré dans l'intron du minigène « mère » grâce à des sites de restriction uniques (Figure 32 ; Cooper, 2005 ; Gaildrat *et al.*, 2010). En l'absence d'ADNg de patient, il est également possible d'introduire la variation par mutagenèse dirigée au sein de la séquence d'intérêt. Cette possibilité est un des avantages majeurs du test fonctionnel basé sur l'utilisation de minigènes, puisqu'il n'est pas dépendant de la disponibilité du matériel biologique du patient. Après la transfection en parallèle des minigènes WT et muté dans des modèles cellulaires en culture, il est alors possible de réaliser une analyse comparative de leurs profils d'épissage par RT-PCR et séquençage (Figure 32 ; Baralle and Baralle, 2005 ; Gaildrat *et al.*, 2010).

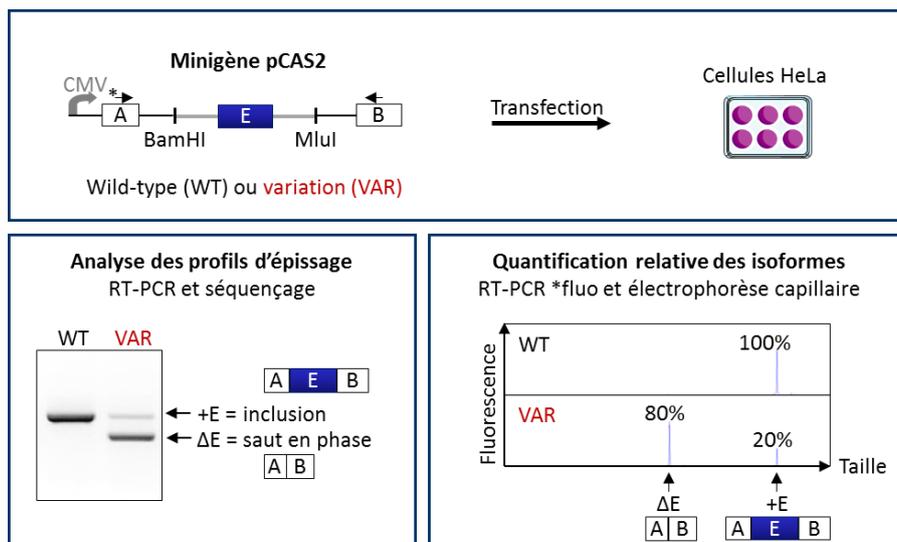


Figure 32. Principe du test fonctionnel indicateur d'anomalies d'épissage, basé sur l'utilisation d'un minigène. L'exemple présenté ici est celui du minigène pCAS2.

Cette technique est particulièrement versatile car *a priori* adaptable à l'analyse de toute altération susceptible de modifier l'épissage, qu'elle soit exonique ou intronique, ponctuelle ou impliquant plusieurs nucléotides, et ce potentiellement dans n'importe quel gène multi-exon. De plus, contrairement à l'analyse de l'ARN des patients, cette approche permet d'établir un lien direct de causalité entre la variation et les événements aberrants d'épissage puisque l'analyse des minigènes sauvage et mutés est réalisée en parallèle dans des contextes cellulaire et nucléotidiques identiques, rendant l'interprétation des données plus aisée (Cooper, 2005). De surcroît, ce système est mono-allélique ce qui facilite l'analyse par comparaison aux analyses de l'ARN des patients, par définition effectuées dans un contexte bi-allélique. En outre, cela n'exclut pas pour autant l'éventualité d'évaluer l'effet combiné sur l'épissage de plusieurs variations localisées sur le même allèle (de la Hoya *et al.*, 2016). Cette technique favorise également l'analyse de transcrits aberrants tels que ceux porteurs d'un PTC. En effet, il est possible d'inhiber la dégradation de ce type de transcrits par le NMD par le biais d'un traitement cellulaire à l'aide d'agents inhibiteurs de la synthèse protéique, tels que la puromycine ou la cycloheximide, ou directement par inactivation du codon d'initiation de la traduction du minigène, comme c'est le cas pour le vecteur pCAS2 développé par notre groupe (Soukariéh *et al.*, 2016). Enfin, il est envisageable d'appréhender l'impact du contexte cellulaire sur les événements aberrants d'épissage induits par une variation en transfectant les minigènes d'intérêt dans des lignées issues de différents tissus (Tournier *et al.*, 2008 ; de la Hoya *et al.*, 2016).

Des minigènes spécifiques à l'analyse des exons terminaux ont également été développés, notamment pour l'étude du premier exon du gène *MSH2* ou encore le dernier exon du gène *CDH1* (Naruse *et al.*, 2009 ; Grodecka *et al.*, 2014b). Récemment, Raud et ses collaborateurs ont mis au point un minigène universel pSplicePOLR2G3 permettant d'analyser les variations localisées dans le premier exon ou le dernier exon d'un gène (Raud *et al.*, 2019). Il est important de noter que les minigènes de petite taille (e.g. à 3 exons) ne permettent pas d'évaluer la totalité des profils d'épissage potentiels, comme par exemple le saut combiné de plusieurs exons ou des isoformes alternatives plus complexes. C'est dans le but d'appréhender l'impact des variations dans un contexte génomique plus large que des constructions plasmidiques avec un plus grand nombre d'exons consécutifs ont été développés (Raponi *et al.*, 2011 ; Acedo *et al.*, 2012, 2015 ; Sharma *et al.*, 2014 ; Ramalho *et al.*, 2015 ; Fraile-Bethencourt *et al.*, 2017, 2018, 2019). On peut notamment citer le minigène pB1 qui a été développé dans le but d'étudier l'épissage alternatif du gène *BRCA1* entre les exons 8 et 12 ainsi que l'impact des variations localisées dans l'exon 11 de ce gène (Raponi *et al.*, 2012), ou encore, le vecteur MGBR2ex19_27 qui constitue le plus large minigène construit pour l'étude des variations identifiées dans *BRCA2* (exons 19 à 27) (Acedo *et al.*, 2015).

Bien que le test indicateur d'anomalies d'épissage basé sur l'utilisation de minigène continue d'être régulièrement amélioré et présente de nombreux avantages, il n'en reste pas moins que cette approche est artificielle étant donné les contextes cellulaires et nucléotidiques utilisés. Malgré les limites associées à cette approche, il est intéressant de constater que la majorité des résultats obtenus dans le contexte des minigènes présentent une très grande concordance avec des données établies à partir du matériel biologique de patients, démontrant ainsi la pertinence de cette approche pour la détection des défauts d'épissage induits par des variations et son utilité pour leur interprétation biologique et clinique (Tournier *et al.*, 2008 ; Gaildrat *et al.*, 2012 ; van der Klift *et al.*, 2015 ; Soukariéh *et al.*, 2016 ; Baralle and Buratti., 2017 ; Tubeuf *et al.*, 2020a). Néanmoins, quelques rares données discordantes obtenues au sein de notre équipe montrent qu'il reste essentiel de valider chaque minigène, surtout pour des raisons de calibration (type et sévérité des défauts d'épissage), par la confirmation des effets observés à partir de l'ARN de patients. Ces approches sont donc complémentaires pour l'interprétation biologique et clinique des variations splicéogéniques.

Le test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes a permis d'évaluer et de caractériser expérimentalement un grand nombre de variations splicéogéniques affectant les sites d'épissage mais également les éléments de régulation de l'épissage, non seulement dans nos gènes d'intérêt, *BRCA* et *MMR*, mais également dans de nombreux autres gènes tels que *CFTR*, *NF1*, *SMN1* et *SMN2* (Pagani *et al.*, 2003 ; Pagani *et al.*, 2005 ; Hernandez-Imaz *et al.*, 2015 ; Singh *et al.*,

2007). Ces données constituent aujourd'hui une aide précieuse pour l'interprétation biologique et clinique des variations dans les gènes associés aux maladies mendéliennes dans le cadre du diagnostic, avec des conséquences importantes pour l'optimisation de la prise en charge des patients et de leurs apparentés. Cependant, une des limites importantes de ces approches est le débit d'analyse qui reste faible puisque chaque variation fait ici l'objet d'une analyse individuelle dans un essai qui requiert au moins 2 semaines de travail expérimental comprenant les étapes de clonage, de transfection et d'analyse des transcrits, même si l'analyse de plusieurs variations peut être effectuée en parallèle. Ces essais ne peuvent alors répondre à la demande croissante d'analyse de la pléthore de variations identifiées dans le génome des patients. Afin de dépasser cette limite majeure, des efforts considérables sont actuellement entrepris pour la mise au point de tests fonctionnels à haut débit. Le second paramètre important d'amélioration de ces tests indicateurs d'anomalies d'épissage est l'utilisation d'un contexte génomique d'analyse qui s'approche au plus près du contexte génomique naturel.

IV.3.2.b. Test indicateur d'anomalies d'épissage basé sur l'utilisation de transgènes ou de modifications du génome

Dans le but d'étudier l'effet des variations sur l'épissage dans un contexte génomique le plus naturel possible, des chercheurs ont récemment développé un nouvel essai d'épissage (*Full-Length Gene Splicing Assay, FLGSA*) basé sur la transfection d'un vecteur d'expression renfermant certains gènes du codon d'initiation de la traduction au codon de terminaison de la traduction (Lin *et al.*, 2019 ; Chen *et al.*, 2020). Cette approche est particulièrement intéressante mais reste aujourd'hui limitée à des gènes présentant une taille relativement restreinte (≤ 8 Kb).

Cette contrainte peut être dépassée par le biais de l'utilisation de vecteurs spécifiques, appelés *BAC (Bacterial Artificial Chromosome)*, dans lesquels il est possible de cloner de larges fragments génomiques (jusqu'à 300 Kb). L'équipe du Dr Sharan a ainsi mis à profit des vecteurs BAC renfermant l'intégralité des gènes humains *BRCA1* ou *BRCA2* (avec leurs promoteurs respectifs) pour étudier l'impact des variations sur l'épissage mais également sur la fonction de la protéine (Figure 33 ; Yang and Sharan, 2003 ; Kuznetsov *et al.*, 2008). Après introduction de ces vecteurs sauvages ou mutés dans des cellules souches de souris (mESC, *mouse Embryonic Stem Cells*), l'impact des variations introduites par mutagenèse est examiné par l'analyse RT-PCR des profils d'épissage des transcrits produits à partir des transgènes BAC (Yang and Sharan, 2003 ; Kuznetsov *et al.*, 2008 ; Biswas *et al.*, 2011, 2012 ; Mesman *et al.*, 2019, 2020). Il est toutefois important de noter qu'initialement cet essai a été principalement utilisé pour étudier l'impact de variations de type faux-sens sur la protéine. En effet,

cette approche permet également d'évaluer la fonctionnalité de la protéine produite dans le contexte muté puisque, après inactivation du gène *Brca* endogène des mESC, la survie cellulaire, l'activité de recombinaison homologue et la sensibilité à certains agents génotoxiques deviennent dépendantes de l'activité de la protéine BRCA humaine produite à partir du BAC (Figure 33).

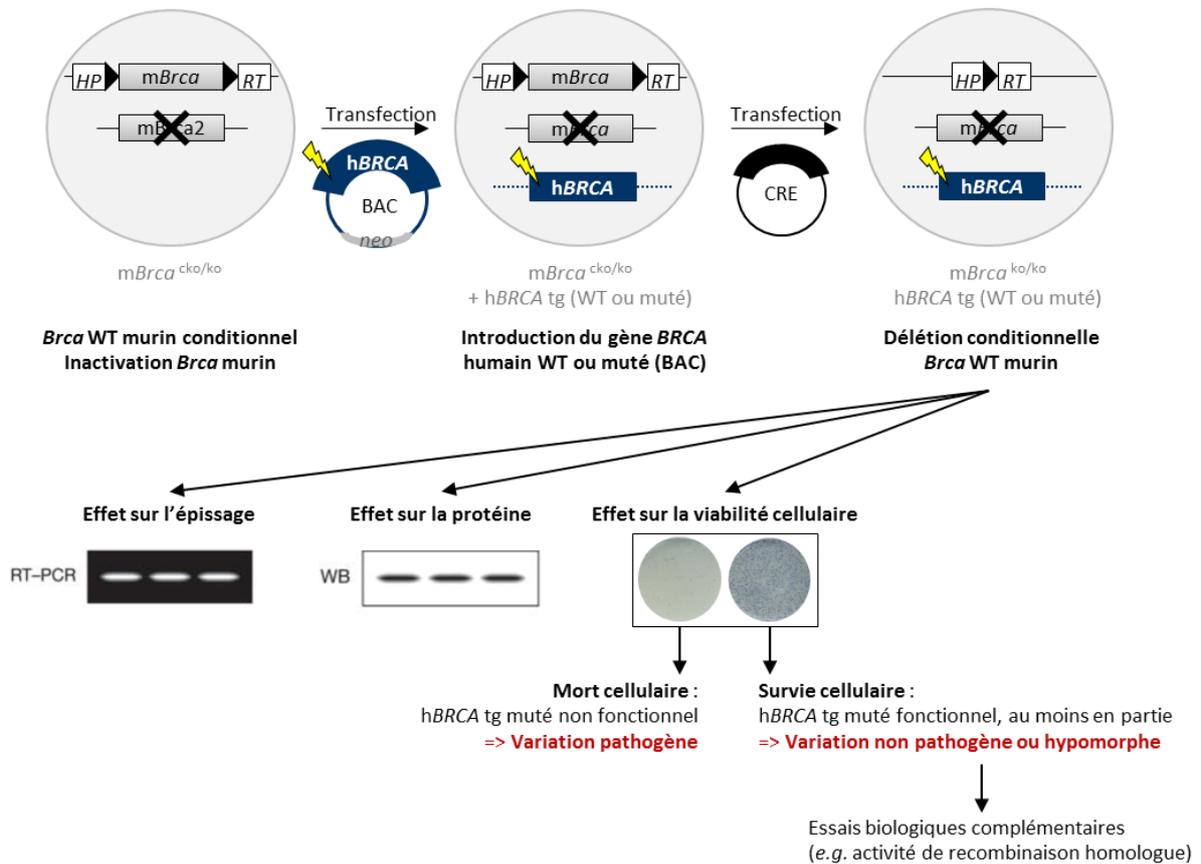


Figure 33. Principe de l'essai fonctionnel basé sur l'utilisation de cellules souches embryonnaires de souris (adapté de Kuznetsov *et al.*, 2008). Désignées sous le terme de *mBRCAcko/ko*, les mESC présentent un allèle conditionnel (*cko*) avec un gène *Brca* murin fonctionnel qui peut être inactivé grâce au système de recombinaison Cre-LoxP (sites LoxP, triangles noirs) et un allèle avec un gène *Brca* murin défectueux (*ko*, *knock out*). Le gène *BRCA* humain entier (*hBRCA* tg, transgène), muté (éclair jaune, variation) ou non (WT, *Wild-type*), est introduit par transfection d'un BAC. L'allèle conditionnel est alors inactivé par transfection de la recombinase Cre.

Ces différentes stratégies permettent de se rapprocher du contexte génomique naturel mais elles requièrent l'expression d'un transgène, avec tous les biais éventuels associés. Afin de dépasser cette limite, il est aujourd'hui possible d'introduire les variations génétiques d'intérêt directement dans le locus chromosomique endogène grâce notamment à une technique révolutionnaire d'ingénierie ciblée du génome (*genome editing*) basée sur l'utilisation des ciseaux moléculaires ou *CRISPR* (*Clustered Regularly Interspaced Short Palindromic Repeats*)-Cas9. Cette remarquable technologie inventée par le Dr Charpentier et le Dr Doudna (Prix Nobel de Chimie 2020) s'impose

aujourd'hui comme la stratégie privilégiée d'édition du génome étant donné sa précision satisfaisante, sa simplicité relative et son faible coût. Dès 2014, l'équipe du Dr Shendure a mis à profit cette technologie pour étudier l'impact, au niveau des transcrits, de toutes les variations possibles localisées dans l'exon 18 du gène *BRCA1* de cellules humaines haploïdes (Findlay *et al.*, 2014). Couplée à l'utilisation d'une librairie complexe de matrices porteuses des mutations, cette approche de saturation par édition du génome a permis d'analyser simultanément l'effet d'un grand nombre de variations introduites dans le génome des cellules transfectées. Ces travaux ont permis d'identifier de très nombreuses variations splicéogéniques responsables de la modification d'ESR (Findlay *et al.*, 2014). De façon tout à fait remarquable, après cette preuve de concept, cette approche a été mise à profit pour l'analyse de près de 4 000 substitutions ponctuelles localisées dans 13 exons fonctionnellement essentiels (exons 2 à 5 et 15 à 23) et aux régions introniques flanquantes de ces exons dans ce même gène, *BRCA1* (Findlay *et al.*, 2018). Des avancées technologiques majeures, telles que (i) la construction multiplexe de bibliothèques de variations génétiques à partir d'ADN synthétique, (ii) l'ingénierie ciblée du génome par l'utilisation de CRISPR-Cas9 et (iii) le séquençage multiplex à haut débit (DNA-seq et RNA-seq), sous-tendent le développement de cet essai. Ainsi, cette approche permet de relier chacune des multiples variations testées dans une seule et unique expérience à un phénotype moléculaire (abondance des transcrits) ou cellulaire (survie) sous la forme d'un score de fonctionnalité individuel (Figure 34 ; Findlay *et al.*, 2018). En effet, l'impact de chacune des milliers de variations testées a été évalué sur la base de l'analyse par RNA-seq de l'abondance des transcrits mais également sur la capacité à produire une protéine BRCA1 fonctionnelle, indispensable à la survie des cellules utilisées dans cet essai. Les données générées dans le cadre de cette étude ont été en partie validées sur la base de la concordance observée ($\geq 96\%$) avec la classification établie pour une série de variations pathogènes et neutres. Des essais fonctionnels comparables, mettant à profit l'introduction des variations dans le gène endogène et l'évaluation d'un phénotype cellulaire associé, ont également été développés récemment pour les gènes *MMR*, notamment *MSH2*, dans des cellules souches embryonnaires humaines (Rath *et al.*, 2019). Mais, à ce jour et à notre connaissance, alors que cette approche présente le potentiel d'appréhender l'impact des variations sur l'épissage, elle a principalement été utilisée pour évaluer l'impact de variations faux-sens sur la protéine MSH2 (Rath *et al.*, 2019).

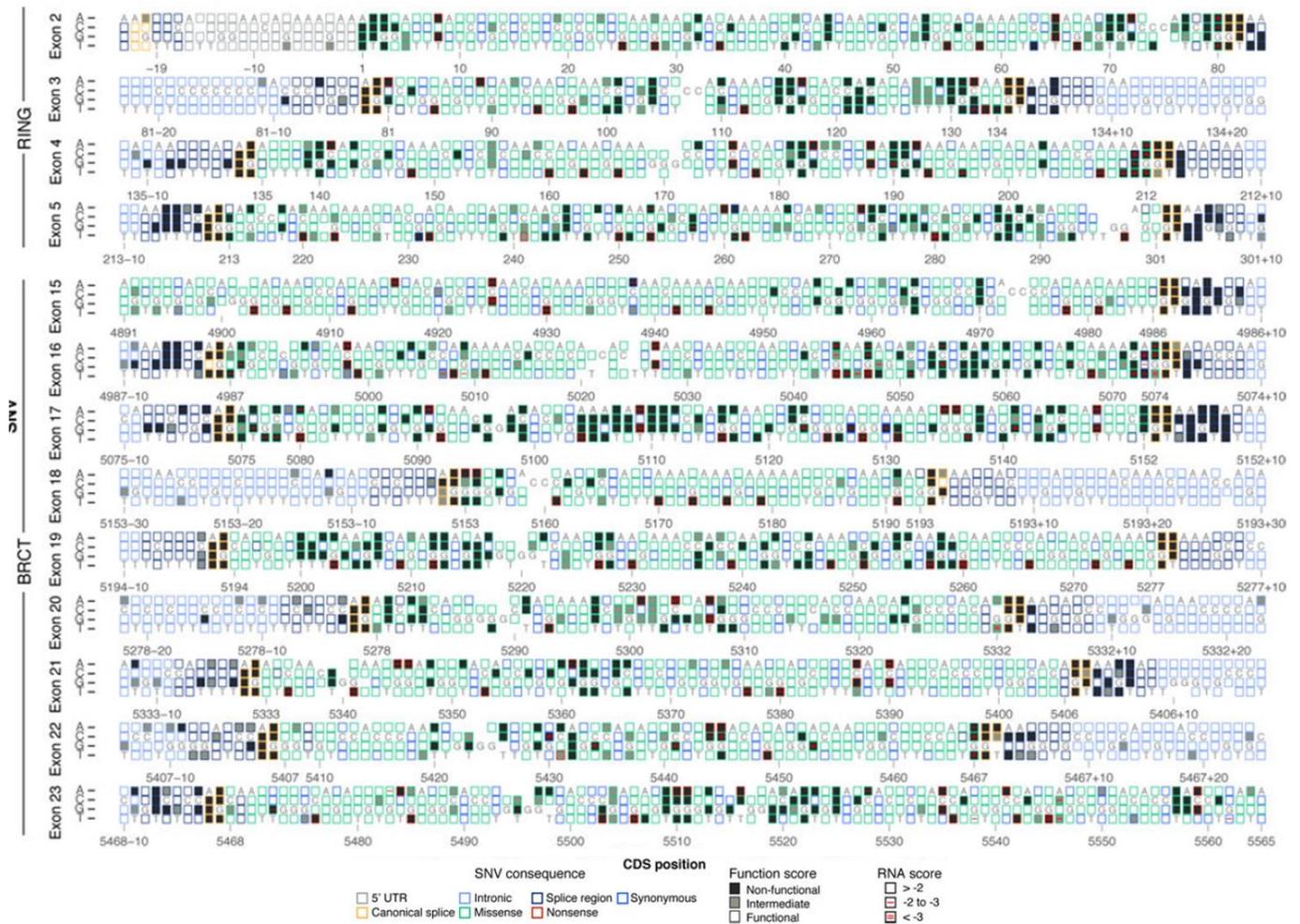


Figure 34. Carte représentant chaque substitution possible dans les exons 2 à 5 et 15 à 23 de *BRCA1* avec leur fonctionnalité associée déterminée par SGE (d'après Findlay *et al.*, 2018).

L'ensemble de ces approches fonctionnelles à plus ou moins haut débit représente aujourd'hui une stratégie puissante et très prometteuse pour l'évaluation de l'impact fonctionnel combiné des VSI sur l'épissage et sur la fonction des protéines. Ces approches représentent également des outils remarquables susceptibles de conduire à une meilleure dissection et compréhension, d'une part, des mécanismes moléculaires sous-tendant l'épissage et, d'autre part, de la relation structure/fonction au niveau de la protéine.

Objectifs des travaux de thèse

Aujourd'hui, suite notamment à l'implémentation des technologies de séquençage à haut débit, le défi majeur en génétique médicale ne réside plus dans la détection des variations présentes dans le génome des patients mais bel et bien dans l'interprétation biologique et clinique de ces altérations. Dans un contexte de maladie monogénique, le but est alors d'identifier la variation génétique causale, à l'origine de la pathologie. Cette problématique est rapidement devenue prioritaire et essentielle pour les laboratoires assurant le diagnostic moléculaire des maladies à déterminisme mendélien car l'identification de la variation pathogène permet d'optimiser la prise en charge médicale des patients et de leurs apparentés, avec des implications en termes de diagnostic, de pronostic, de prévention mais également dans les choix thérapeutiques. En général, les variations présumées responsables de la perte de fonction d'un gène sont systématiquement classées pathogènes. Il s'agit des variations perçues comme introduisant un codon stop prématuré dans la séquence codante ; c'est-à-dire des variations de type non-sens ou de type insertions/délétions à l'origine d'un décalage du cadre de lecture mais également des variations introniques localisées au niveau des dinucléotides introniques les plus conservés des sites d'épissage ($IVS\pm 1/2$). Ces dernières années, il a été montré que de nombreuses variations sont pathogènes du fait de l'altération de l'épissage, processus indispensable à la maturation des ARN pré-messagers. Dans la majorité des cas, ces variations induisent des anomalies résultant en un décalage du cadre de lecture qui entraîne l'introduction d'un codon stop prématuré et une perte de fonction. Toutefois, certaines variations splicéogéniques pourraient être à l'origine de modifications de l'épissage en phase et ainsi, être responsables de la production d'une protéine potentiellement fonctionnelle.

Dans le cadre de mon projet de thèse, nous avons fait l'hypothèse selon laquelle certaines variations $IVS\pm 1/2$ et non-sens, supposées pathogènes dans le contexte de formes héréditaires de cancers, pourraient permettre le maintien d'une fonction, au moins en partie, par le biais d'un tel mécanisme de « sauvetage ». L'objectif principal de mes travaux était de tester cette hypothèse avec comme modèles d'étude deux gènes majeurs de prédispositions aux cancers : *BRCA2* pour le syndrome sein-ovaire et *MSH2* pour le syndrome de Lynch. La stratégie expérimentale envisagée devait mettre à profit des approches complémentaires permettant d'appréhender l'impact des variations non seulement au niveau de l'épissage de l'ARN, mais également au niveau de la fonction de la protéine.

Dans le cadre de mes travaux portant sur le gène *BRCA2*, nous avons ciblé l'exon 12 pour deux raisons principales : d'une part, cet exon est en phase et, d'autre part, il avait déjà été décrit comme un exon non essentiel à la fonction de recombinaison homologue de la protéine BRCA2 (Li *et al.*, 2009). Il s'agissait alors :

1. d'évaluer la splicéogénicité de toutes les variations IVS±1/2 et non-sens répertoriées, au niveau de cet exon, dans les bases de données de patients. Les approches expérimentales mises en œuvre pour la réalisation de cette étape ont été le test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes ainsi que l'analyse des profils d'épissage et du déséquilibre allélique à partir de l'ARN de patients.
2. de déterminer les conséquences fonctionnelles au niveau de la protéine BRCA2 des variations splicéogéniques caractérisées lors de la première étape du projet. Cette seconde étape a mis à profit un essai basé sur l'utilisation de cellules souches embryonnaires de souris (*i.e.* test de complémentation, mesure de l'activité de recombinaison homologue, sensibilité aux inhibiteurs de PARP).

Mes travaux portant sur le gène *MSH2* avaient pour objectif de caractériser d'un point de vue fonctionnel les différents biotypes de défauts d'épissage en phase potentiellement induits par des variations IVS±1/2. Il s'agissait alors :

1. d'évaluer l'impact sur l'épissage de certaines variations IVS±1/2 répertoriées dans les bases de données de patients soit au niveau de chaque exon en phase de *MSH2*, soit au niveau d'exons hors phase de ce gène, mais potentiellement responsables de délétions exoniques ou rétentions introniques partielles. L'approche expérimentale choisie pour cette analyse a été le test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes.
2. de déterminer les conséquences fonctionnelles au niveau de la protéine MSH2 des modifications en phase provoquées par les IVS±1/2 splicéogéniques caractérisées dans la première étape de ce projet. La méthode retenue pour cette étude a été un test basé sur la tolérance à la méthylation.

L'ensemble de ces travaux devait permettre (i) de vérifier l'existence potentielle d'un mécanisme de « sauvetage » de certaines variations présumées pathogènes, par le biais d'une altération en phase de l'épissage, dans le contexte des formes héréditaires de cancers et (ii) de questionner alors la classification pathogène de ce type de mutations, avec des conséquences potentiellement importantes pour la prise en charge médicale des patients et de leurs apparentés.

Résultats

I. Toutes les variations non-sens ne conduisent pas à une perte totale de fonction : le paradigme de l'exon 12 de *BRCA2*

L'exon 12 de *BRCA2* représentait un excellent modèle pour tester notre hypothèse selon laquelle certaines des variations supposées pathogènes, telles que les variations au niveau des sites canoniques d'épissage (IVS±1/2) et les mutations non-sens, dans les gènes de prédisposition aux cancers pourraient ne pas conduire à une perte totale de fonction du fait de leur impact sur l'épissage. Il a précédemment été montré que la séquence codée par cet exon en phase était non essentielle à la fonctionnalité de la protéine BRCA2 (Li *et al.*, 2009). La mise en évidence de la redondance fonctionnelle de cet exon s'est initialement appuyée sur l'étude d'une variation faux-sens (c.6853A>G, p.Ile2285Val), à l'origine du saut de cet exon et qui a été classée neutre sur la base de données de co-ségrégation dans les familles de patients et de co-occurrence en *trans* avec une autre variation pathogène (Easton *et al.*, 2007 ; Tavgitigian *et al.*, 2008 ; Li *et al.*, 2009).

Notre stratégie a donc consisté à tester l'impact des variations IVS±1/2 et non-sens sur l'épissage de cet exon, à l'aide d'approches complémentaires basées sur l'utilisation de minigènes et d'analyses à partir de lignées lymphoblastoïdes de patientes porteuses de ces variations. Nos travaux ont ainsi permis de démontrer que les IVS±1/2 mais également certaines variations non-sens sont responsables du saut en phase de l'exon 12, du fait de l'altération des sites 5' et 3' d'épissage ou d'ESRs. Sur la base de ces résultats, nous avons sélectionné certaines de ces variations pour évaluer leurs conséquences fonctionnelles au niveau de la protéine BRCA2. Dans le cadre d'une collaboration avec l'équipe du Dr Vreeswijk (Centre Médical Universitaire, Leiden, Pays-Bas), nous avons pu confirmer les effets de ces variations dans le contexte du gène humain entier introduit dans des cellules souches embryonnaires modifiées de souris. Après inactivation du gène *Brca2* de souris dans ce modèle cellulaire, l'évaluation de la survie cellulaire et de la capacité de réparation des cassures double brin de l'ADN par recombinaison homologue a révélé que les variations IVS±1/2 et non-sens splicéogéniques étaient bien à l'origine d'une activité BRCA2 partielle, résultant de l'expression de la protéine dépourvue de l'exon 12.

L'ensemble de ces données a permis d'apporter des arguments en faveur de l'existence d'un mécanisme de sauvetage de certaines variations supposées pathogènes, du fait de leur impact direct sur l'épissage conduisant au saut en phase d'une séquence non-essentielle à la fonction de recombinaison homologue de la protéine BRCA2. Même si d'autres fonctions sont associées à BRCA2, l'activité de recombinaison homologue apparaît comme déterminante dans le rôle de cette protéine en tant que suppresseur de tumeur. C'est en effet la déficience de cette fonction spécifique qui semble

être capitale dans le processus d'oncogénèse observé dans les tumeurs de patients porteurs de variations pathogènes de *BRCA2*. La persistance d'une activité de recombinaison homologue intermédiaire (30 à 50%) ou similaire à celle associée à des variations neutres (>50%) (Mesman *et al.*, 2018) observée dans le contexte de variations IVS±1/2 et non-sens de l'exon 12 de *BRCA2* remet en cause l'interprétation biologique de ces variations initialement supposées responsables d'une perte totale de fonction. Sur la base de ces données, la classification clinique pathogène initialement assignée à ces variations a été remise en question. Afin d'appréhender le lien entre ces variations et le risque de développer un cancer, les données cliniques et génétiques disponibles pour les familles de patientes porteuses de ces variations spécifiques ont été collectées, en collaboration avec le réseau français des laboratoires d'oncogénétique. Si plusieurs observations peuvent être conformes avec le caractère hypomorphique de ces variations, ces données restent insuffisantes pour statuer sur leur classification clinique. Des études internationales de co-ségrégation et de cas/témoins à plus grande échelle devraient permettre d'estimer le risque de cancers associé à ces variations hypomorphes. En prenant en considération l'effet biologique de ces variations, nous proposons pour l'instant de reclasser en variations de signification clinique inconnue (classe 3), en accord avec les recommandations émises par les consortia ACMG/AMP et ENIGMA.

Cette étude menée dans le cadre de ma thèse a donc permis de démontrer, pour la première fois dans un gène de prédisposition aux cancers, que la modification directe du processus d'épissage induite par certaines variations est susceptible de contourner la perte totale de fonction attribuée *a priori* à des variations nulles de type non-sens. L'existence d'isoformes en phase potentiellement responsables d'une activité partielle doit désormais être pris en compte pour l'interprétation biologique et clinique de certaines variations de *BRCA2*. Cette découverte souligne l'importance des analyses fonctionnelles combinées qui appréhendent l'impact des variations sur l'ARN et la protéine.

L'ensemble de ces travaux a fait l'objet d'un article publié dans le journal *Cancer Research* (Meulemans *et al.*, 2020 ; 80:1374–1386).

Skipping Nonsense to Maintain Function: The Paradigm of *BRCA2* Exon 12



Laëtitia Meulemans¹, Romy L.S. Mesman², Sandrine M. Caputo^{3,4}, Sophie Krieger^{1,5,6}, Marine Guillaud-Bataille⁷, Virginie Caux-Moncoutier^{3,4}, Mélanie Léone⁸, Nadia Boutry-Kryza⁸, Johanna Sokolowska⁹, Françoise Révillion¹⁰, Capucine Delnatte¹¹, Hélène Tubeuf^{1,12}, Omar Soukarieh¹, Françoise Bonnet-Dorion¹³, Virginie Guibert¹¹, Myriam Bronner⁹, Violaine Bourdon¹⁴, Sarab Lizard⁹, Paul Vilquin¹⁵, Maud Privat^{16,17}, Aurélie Drouet¹, Charlotte Grout¹, Fabienne M.G.R. Calléja², Lisa Golmard^{3,4}, Harry Vrieling², Dominique Stoppa-Lyonnet^{3,18}, Claude Houdayer^{1,3,19}, Thierry Frebourg^{1,19}, Maaik P.G. Vreeswijk², Alexandra Martins¹, and Pascaline Gaildrat¹

ABSTRACT

Germline nonsense and canonical splice site variants identified in disease-causing genes are generally considered as loss-of-function (LoF) alleles and classified as pathogenic. However, a fraction of such variants could maintain function through their impact on RNA splicing. To test this hypothesis, we used the alternatively spliced *BRCA2* exon 12 (E12) as a model system because its in-frame skipping leads to a potentially functional protein. All E12 variants corresponding to putative LoF variants or predicted to alter splicing ($n = 40$) were selected from human variation databases and characterized for their impact on splicing in minigene assays and, when available, in patient lymphoblastoid cell lines. Moreover, a selection of variants was analyzed in a mouse embryonic stem cell-based functional assay. Using these complementary approaches, we demonstrate that a subset of variants, including nonsense variants,

induced in-frame E12 skipping through the modification of splice sites or regulatory elements and, consequently, led to an internally deleted but partially functional protein. These data provide evidence, for the first time in a cancer-predisposition gene, that certain presumed null variants can retain function due to their impact on splicing. Further studies are required to estimate cancer risk associated with these hypomorphic variants. More generally, our findings highlight the need to exercise caution in the interpretation of putative LoF variants susceptible to induce in-frame splicing modifications.

Significance: This study presents evidence that certain presumed loss-of-function variants in a cancer predisposition gene can retain function due to their direct impact on RNA splicing.

Introduction

Germline nonsense variants account for 20% of all Mendelian disease-causing single-nucleotide variants (SNV) reported within gene coding sequences (1). This type of variants, as well as frameshift deletions/insertions and SNVs located at the highly conserved intronic dinucleotides of splice sites (hereafter termed $IVS\pm 1/2$), are presumed to lead to loss-of-function (LoF) because they are expected to introduce a premature termination codon (PTC). On the basis of these assumptions, these variants are generally regarded as null alleles and classified as pathogenic without further investigation. This is in particular the case for the iconic cancer predisposition genes, *BRCA1* and *BRCA2*, where mono-allelic germline dominant LoF variants confer high risks of hereditary breast and ovarian cancer (HBOC; ref. 2).

Several consortia that provide guidelines for variant classification in the clinical setting, including the American College of Medical Genetics and Genomics (ACMG; ref. 3) and the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA; refs. 4, 5), recommend caution with the systematic pathogenic classification of such presumed null variants. Indeed, possible rescue mechanisms have been described for (partial) function preservation of certain presumed LoF variants in *BRCA1* and *BRCA2* genes. In particular, the polymorphic *BRCA2* c.9976A>T (p.Lys3326^{*}) nonsense variant results in a truncated protein with a minor loss in the C-terminal region that preserves functionality (6–8) and is associated with a modest increase in HBOC risks compared with typical LoF pathogenic *BRCA2* variants (9). A second rescue mechanism involves specific naturally occurring alternative splicing events, as recently illustrated by the characterization of *BRCA1* c.[594-2A>C; 641A>G] (5). This allele is

¹Normandie Univ, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen, France. ²Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. ³Department of Genetics, Institut Curie, Paris, France. ⁴PSL Research University, Paris, France. ⁵Laboratory of Cancer Biology and Genetics, Centre François Baclesse, Caen, France. ⁶Normandie University, UNICAEN, Caen, France. ⁷Department of Genetics, Institut Gustave Roussy, Villejuif, France. ⁸Hospices Civils de Lyon, Bron, France. ⁹Department of Genetics, Nancy University Hospital, Nancy, France. ¹⁰Unit of Human Molecular Oncology, Centre Oscar Lambret, Lille, France. ¹¹Department of Genetics, Nantes University Hospital, Nantes, France. ¹²Interactive Biosoftware, Rouen, France. ¹³Inserm U916, Department of Genetics, Institut Bergonié, Bordeaux, France. ¹⁴Department of Genetics, Institut Paoli-Calmettes, Marseille, France. ¹⁵Department of Pathology and Oncobiology, Montpellier University Hospital, Montpellier, France. ¹⁶University of Clermont Auvergne, Inserm U1240, Clermont Ferrand, France. ¹⁷Department of Oncogenetics, Centre Jean Perrin, Clermont Ferrand, France. ¹⁸Inserm U830, University Paris Descartes, Paris, France. ¹⁹Department of Genetics, Rouen University Hospital, Rouen, France.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

L. Meulemans and R.L.S. Mesman contributed equally to this article.

Corresponding Author: Pascaline Gaildrat, Normandie Univ, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen 76183, France. Phone: 332-3514-8315; E-mail: pascaline.gaildrat@univ-rouen.fr

Cancer Res 2020;80:1374–86

doi: 10.1158/0008-5472.CCR-19-2491

©2020 American Association for Cancer Research.

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

not associated with high cancer risk despite its induced out-of-frame exon 10 skipping, potentially because the naturally occurring in-frame alternative transcript without exons 9 and 10 ($\Delta E9_E10$) can encode for a functioning BRCA1 protein. Naturally occurring alternative splicing can also enable survival of patients homozygous for certain BRCA1 exon 11 nonsense variants, albeit with severe features of Fanconi anemia (10). Here, we raised the hypothesis that a subset of presumed LoF variants in HBOC susceptibility genes may potentially retain partial activity through their direct impact on splicing, leading eventually to questioning their clinical interpretation.

Virtually all variants may potentially induce a perturbation in RNA splicing, notably through the modification of either 3' and 5' splice sites (3'/5' ss) or exonic splicing regulatory elements (ESR; refs. 11, 12). The most frequently observed splicing anomaly is exon skipping that often leads to a frameshift, causing LoF. However, some specific presumed LoF variants could induce in-frame skipping of a nonessential exon, resulting in a functional protein, at least partially. Here, we envisioned that this rescue mechanism may be triggered by specific variants located in the exon 12 (E12) of BRCA2. This exon does not encode a known functional domain involved in the BRCA2 tumor suppressor activity and its in-frame skipping leads to the internal deletion of a potentially dispensable 32-amino acid sequence (13). In a mouse embryonic stem cell (mESC)-based functional assay, the E12-deleted BRCA2 protein is able to complement the loss of endogenous wild-type *Brca2* and exhibits the same functions with regard to (i) homologous recombination activity, (ii) radiation induced RAD51 foci formation, and (iii) sensitivity to DNA-damaging agents (13).

An additional argument supporting the tumor suppressor functionality of the BRCA2 E12-deleted protein came from the initial analysis of the clinical significance of a missense variant, c.6853A>G (p.Ile2285Val), that triggers E12 skipping. The family history likelihood ratio and the cooccurrence *in trans* with a pathogenic variant support the classification of this spliceogenic variant as neutral (13).

In this study, we selected, from human variation databases, a total of 40 variants in or near BRCA2 E12 that were either presumed LoF variants or predicted to alter E12 splicing, and characterized their impact on splicing in minigene assays and patient biological material. Moreover, a selection of variants was tested in a mESC-based functional assay to assess their combined effects on RNA splicing and protein activity. To evaluate their association with cancer, clinical and family data were collected for spliceogenic BRCA2 E12 variants. Altogether, these data provide evidence that presumed LoF variants, including nonsense SNVs, can induce splicing alterations that trigger in-frame deletion of a sequence not essential for BRCA2 function. The existence of this potential rescue mechanism instigates the need to reevaluate the interpretation of certain BRCA2 variations currently considered as LoF.

Materials and Methods**Variant selection**

All variants located in BRCA2 E12 and its flanking intronic sequences were extracted from human variation databases (Supplementary Fig. S1). First, we selected the SNVs generally regarded as LoF variants, including: (i) $IVS\pm 1/2$, (ii) nonsense, and (iii) frameshift variants. Second, we extended this selection to all variants predicted to affect E12 splicing based on two types of bioinformatics approaches. Predictions of the variant effects on 3'/5' ss were obtained by using the MaxEntScan (MES) *in silico* tool (14), interrogated from the integrated software Alamut Visual 2.9 (Interactive Biosoftware). Predictions of variant impacts on ESR were generated by using a bioinformatics

approach relying on $\Delta tESRseq$ values (total ESRseq score changes; refs. 15–17), based on exonic hexamer scores. Variants that were associated with either a decrease by at least 15% of MES score ($\Delta MES \leq -15\%$) or a $\Delta tESRseq$ score inferior to -0.7 ($\Delta tESRseq < -0.7$) were retained for splicing minigene analysis.

Splicing minigene assay

To evaluate the impact of the selected variants on E12 splicing, we performed an *in cellulo* splicing assay based on the use of the two-exon minigene pCAS2 vector (17, 18). Wild-type (WT) and mutant segments encompassing BRCA2 E12 and its intronic flanking sequences were amplified from patient genomic DNA (gDNA) by PCR using specific primers (Supplementary Table S1) and inserted into the intron of the pCAS2 minigene using *Bam*HI and *Mlu*I restriction sites, yielding pCAS2-BRCA2-E12 minigenes (Fig. 1A). When patient gDNA was not available, the variant of interest was introduced by site-directed mutagenesis. WT and mutant pCAS2-BRCA2-E12 minigenes were transfected in parallel into HeLa cells using the FuGENE 6 Transfection Reagent (Roche Applied Science). Total RNA was extracted 24 hours posttransfection using the NucleoSpin RNA Kit (Macherey-Nagel). The splicing patterns were analyzed by semiquantitative fluorescence RT-PCR (25 cycles) by using the OneStep RT-PCR Kit (Qiagen), in a 25 μ L reaction volume, with 200 ng total RNA and specific minigene primers (Fig. 1A; Supplementary Table S1). RT-PCR products were separated by electrophoresis on 2% agarose gels. All RT-PCR DNA bands were gel-purified and sequenced. Because of the formation of heteroduplexes between the different RT-PCR products, the images of the gels were not used for quantification. For semiquantitative analysis, fluorescence RT-PCR products were resolved under denaturing conditions by capillary electrophoresis on an automated 3500 Genetic Analyzer (Applied Biosystems). Electropherograms were then analyzed by using the GeneMapper v5.0 Software (Applied Biosystems) and peak areas were used to quantify the relative levels of each transcript.

Analysis of E12 splicing patterns and allele-specific expression in lymphoblastoid cell lines

The EBV-immortalized lymphoblastoid cell lines (LCL) of healthy donors and patients were collected in collaboration with French HBOC diagnostic laboratories within the Group "Genetic and Cancer" (GGC, Unicancer; Supplementary Data). These LCLs were cultured and treated or mock-treated with puromycin as described previously (17). After extraction of total RNA, the E12 splicing patterns were analyzed by semiquantitative fluorescence RT-PCR (26 cycles), as described above, with primers mapping to BRCA2 exons 11 and 14 (Supplementary Table S1).

To measure allele-specific expression (ASE) of E12-containing transcripts (+E12), a SNaPshot quantitative primer extension assay was performed as described previously (17), with the same RT-PCR conditions as described above and using specific primers targeting the variants (Supplementary Table S1). The ASE_{variant} was then calculated after normalization of the cDNA ratio to the corresponding gDNA ratio.

To extend these ASE analyses to both transcripts with (+E12) and without ($\Delta E12$) E12, we set up a SNaPshot protocol targeting the c.7242A>G (p.Ser2414 =) polymorphism located in BRCA2 exon 14 (SNP rs1799955, minor allele frequency in gnomAD = 22.5%). Transcripts +E12 and $\Delta E12$ were selectively amplified in parallel RT-PCR reactions (32 cycles) using a forward primer overlapping either exons 11/12 or 11/13, respectively, combined with a reverse primer in exon 15 (Supplementary Table S1). The SNaPshot

Meulemans et al.

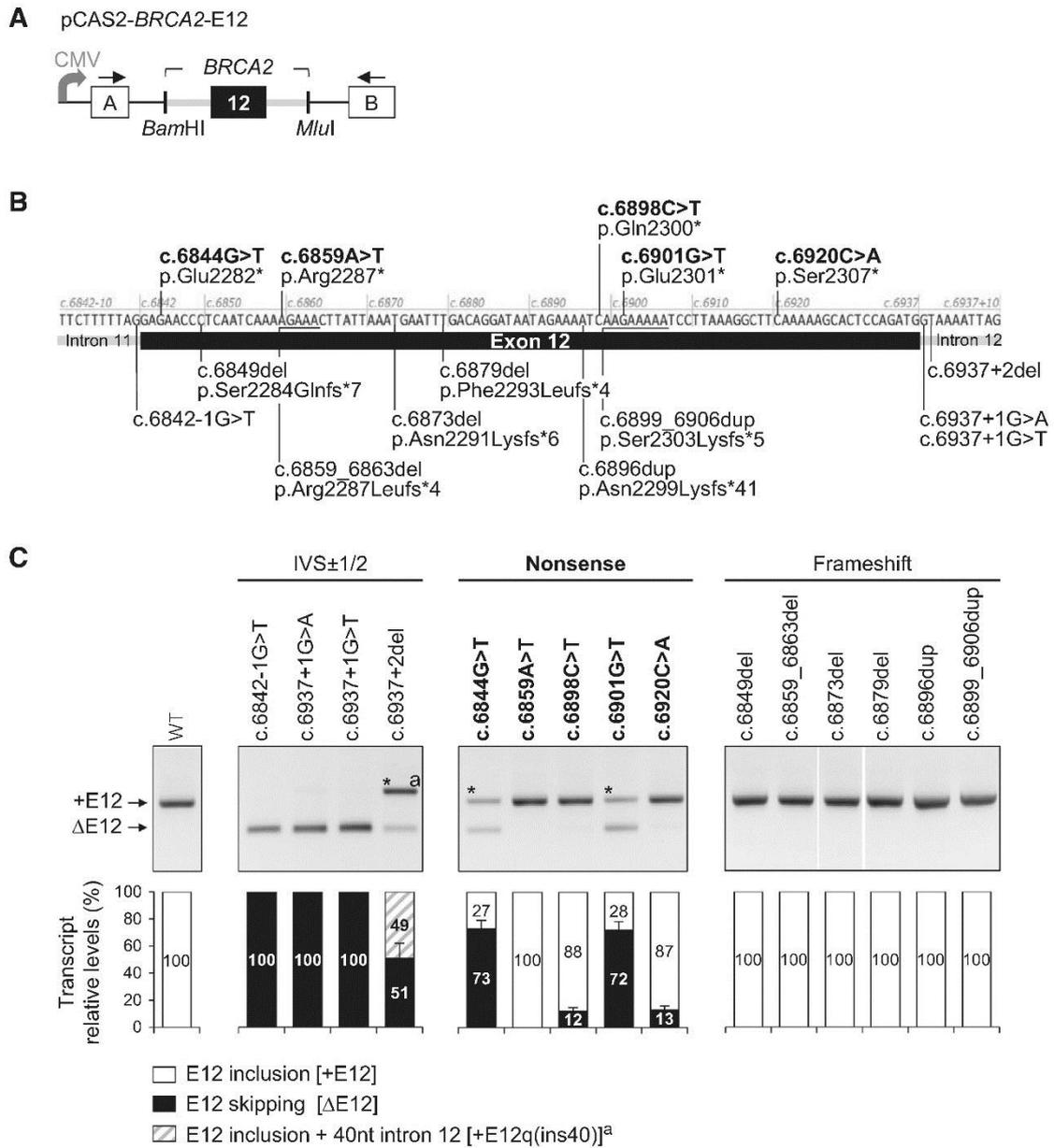


Figure 1.

Identification in minigene assays of *BRCA2* exon 12 spliceogenic variants among presumed LoF SNVs. **A**, Schematic representation of the pCAS2-*BRCA2*-E12 minigene used in the splicing reporter assay. Boxes indicate exons and lines between represent introns. The pCAS2-*BRCA2*-E12 minigenes were generated by inserting a genomic fragment encompassing *BRCA2* exon 12 into the intron of pCAS2 by using *Bam*HI and *Mlu*I restriction sites. Arrows above the exons A and B indicate the position of primers used in RT-PCR analysis. CMV, CMV promoter. **B**, Distribution of 15 presumed LoF variants in *BRCA2* exon 12. The figure shows the nucleotide sequence of *BRCA2* exon 12 and part of its flanking intronic regions, as well as the positions of the 15 selected variants, including four intronic variants located at position IVS±1/2 (below the sequence), five nonsense (above the sequence, in bold), and six frameshift (below the sequence) variants. **C**, Impact of the 15 presumed LoF variants on *BRCA2* exon 12 splicing. WT and mutant pCAS2-*BRCA2*-E12 minigene constructs were transiently expressed in HeLa cells. The splicing patterns of the RNA produced from the different minigenes were analyzed by RT-PCR as described in Materials and Methods. Images show agarose gel electrophoresis of the RT-PCR products. The identities of the two major RT-PCR products, with (+E12) or without (ΔE12) exon 12, are indicated on the left of the gel. The other isoform identified (a) corresponds to the inclusion of exon 12 plus the retention of the first 40 nucleotides of intron 12 [+E12q (ins40)]. Here, it is important to note that the gel images were not used for the quantification of the relative transcript levels because of the formation of heteroduplex between the different RNA isoforms (noted by the star symbol on the gel images), as revealed by band sequencing. Instead, the relative quantifications of the different RNA isoforms were determined by capillary electrophoresis of the fluorescence RT-PCR products (see Supplementary Fig. S2), as indicated in Materials and Methods. The data shown represent the average of at least three independent experiments. Error bars, SD values. nt, nucleotide.

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

quantitative primer extension assay was then performed using a primer targeting the sequence immediately upstream the SNP. The ASE_{SNP} was then calculated as described above.

mESC-based assay

The mESC-based assay (8, 19) was used to assess the functional impact of *BRCA2* E12 variants. The mESC expressing the human *BRCA2* gene (*hBRCA2*) with specific variants were generated as described previously (20). For each transfection round a new ampule of the parental mES cell line was used and cell lines were checked for the presence of *Mycoplasma* monthly. After Cre-mediated deletion of the conditional mouse *Brca2* allele, cell survival, BRCA2 protein expression, and homology-directed repair (HDR) activity were assessed as described previously (20). All experiments were performed in a 5 weeks timespan. For the drug sensitivity assays, mESC were seeded in quadruplicate at 10,000 cells per well on gelatin-coated 96-well plates to measure the sensitivity of *BRCA2* variants to the DNA-damaging agent cisplatin (Accord Healthcare) and the PARP inhibitor, talazoparib (Axon Medchem). Cells were treated with the following concentrations of cisplatin: 0, 0.04, 0.08, 0.16, 0.31, and 0.63 $\mu\text{mol/L}$ (in 1% PBS). Applied concentrations of talazoparib were 0, 0.08, 0.16, 0.31, 0.63, and 1.25 nmol/L (in 1% DMSO). Viable cells were counted using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.) after 48 hours of treatment followed by 24 hours of additional culturing in normal culture medium without a DNA-damaging compound. The E12 splicing patterns were analyzed by fluorescence RT-PCR (23 cycles) as described above for LCLs.

Patient and family data

Clinical, cosegregation, and family data for patients carrying specific *BRCA2* E12 variants were collected in collaboration with the French GGC network and within the framework of the COsegregation of VARIants in the *BRCA1/2* and *PALB2* genes (COVAR) clinical trial (<https://clinicaltrials.gov/ct2/show/NCT01689584>; Supplementary Data). These variants were previously detected in probands and in relatives undergoing genetic counseling. The criteria for diagnostic variant screening of the *BRCA2* gene were applied according to the current French GGC recommendations and the French National Cancer Institute (INCa) guidelines. Informed written consent for genetic analysis was obtained from all patients.

Results**Variant selection**

From a total of 133 *BRCA2* E12 variants collected from human variation databases, we selected all SNVs generally regarded as LoF variants ($n = 15$) encompassing four IVS \pm 1/2, five nonsense, and six frameshift variants (Supplementary Fig. S1; Supplementary Tables S2 and S3). In addition, all other intronic, missense, or synonymous variants predicted to affect either 3'/5' ss ($\Delta\text{MES} \leq -15\%$, $n = 10$) or ESR ($\Delta\text{tESRseq} < -0.7$, $n = 14$; Supplementary Fig. S1; Supplementary Tables S2 and S3) were retained for experimental analysis, as well as the previously described spliceogenic c.6853A>G variant (13). Altogether, this collection comprised 40 variants (Supplementary Fig. S1; Supplementary Tables S2 and S3) that were assessed for their impact on E12 splicing by using an *in cellulo* minigene assay.

Splicing minigene analysis of presumed LoF variants

In this assay, the WT pCAS2-*BRCA2*-E12 minigene only produced +E12 transcripts, whereas, as expected, the four IVS \pm 1/2 variants had a drastic impact on splicing (Fig. 1B and C; Supplementary Fig. S2). Indeed, c.6842-1G>T, c.6937+1G>A, and c.6937+1G>T induced full

E12 skipping (Fig. 1C; Supplementary Fig. S2), in agreement with the predicted destruction of the 3' or 5' ss (Supplementary Fig. S3). Surprisingly, despite bioinformatics predictions strictly identical to c.6937+1G>A and c.6937+1G>T, c.6937+2del generated, in addition to the transcript lacking E12 ($\Delta\text{E12} = 51\%$), one RNA isoform with E12 plus retention of the 40 first nucleotides of intron 12 [+E12q (ins40) = 49%; Fig. 1C; Supplementary Fig. S2]. This latter isoform results from the activation of a cryptic intronic 5' ss predicted for all three variants but only used in the context of c.6937+2del (Supplementary Fig. S3).

Remarkably, in minigene assays, four of the five selected nonsense variants affected splicing (Fig. 1B and C; Supplementary Fig. S2). More precisely, c.6898C>T and c.6920C>A induced a weak exon skipping ($\Delta\text{E12} = 12\%$ and 13%), whereas c.6844G>T and c.6901G>T led to a major exon skipping ($\Delta\text{E12} = 73\%$ and 72%; Fig. 1C; Supplementary Fig. S2). In contrast, none of the six frameshift variants affected E12 splicing (Fig. 1B and C; Supplementary Fig. S2).

Splicing minigene analysis of variants with a predicted splicing effect

We extended our analysis to all SNVs located outside IVS \pm 1/2 and predicted to alter the E12 3'/5' ss ($\Delta\text{MES} \leq -15\%$) (Fig. 2A; Supplementary Fig. S4). All 10 variants from this group had a splicing impact in the minigene assay, albeit to different extents (Fig. 2B; Supplementary Fig. S5A; Supplementary Table S3). Of note, additional transcripts deleted for either the first [+E12p(del3)] or the last [+E12q(del3)] three nucleotides of E12 were detected in the context of c.6842G>A and c.6935A>G, respectively (Fig. 2B; Supplementary S5A). Finally, the impact on E12 splicing of the variants predicted to modify ESR ($\Delta\text{tESRseq} < -0.7$) was experimentally assessed. In this study, *BRCA2* c.6853A>G was used as a positive spliceogenic control because this variant has been previously shown to promote E12 exclusion in patient LCLs, most likely through the alteration of ESR (13). In the minigene assay, we confirmed that this variant induces E12 skipping, albeit moderately ($\Delta\text{E12} = 16\%$; Fig. 2C; Supplementary Fig. S5B; Supplementary Table S3). Of note, the $\Delta\text{tESRseq}$ score associated with this variant ($\Delta\text{tESRseq} = -0.3$) did not correctly predict its splicing impact when taking into account a $\Delta\text{tESRseq}$ value inferior to -0.7 as a positive indicator. In parallel to this variant, we assessed in the minigene assay the splicing effect of 14 E12 variants predicted to affect ESR ($\Delta\text{tESRseq} < -0.7$; Fig. 2A and C; Supplementary Fig. S5B; Supplementary Table S3). In comparison, among these selected variants, two variants induced a major or near-total exon skipping ($\Delta\text{E12} = 58\%$ and 95%), whereas seven variants exhibited a minor or moderate effect ($3\% \leq \Delta\text{E12} \leq 17\%$; Fig. 2C; Supplementary Fig. S5B; Supplementary Table S3). The remaining six selected variants did not cause splicing defect despite predictions of ESR alteration ($\Delta\text{tESRseq} < -0.7$). Altogether, these data not only confirmed initial findings but, notably, uncovered new spliceogenic variants potentially altering ESR. They also showed that the approach relying on $\Delta\text{tESRseq}$ score is not fully reliable for predicting variant-induced E12 splicing effects.

BRCA2 E12 splicing patterns in patient LCLs

In LCLs of three control individuals, we detected an alternative splicing pattern of *BRCA2* E12 (average $\Delta\text{E12} = 10.3\%$; Fig. 3A and B; Supplementary Fig. S6) in agreement with the published data (21, 22). In a LCL derived from a patient heterozygous for a genomic deletion of *BRCA2* E12 (DelE12), a major increase in the relative levels of ΔE12 transcripts ($\Delta\text{E12} = 64\%$) was observed (Fig. 3B; Supplementary Fig. S6), as expected. A comparable level of exon exclusion (ΔE12

Meulemans et al.

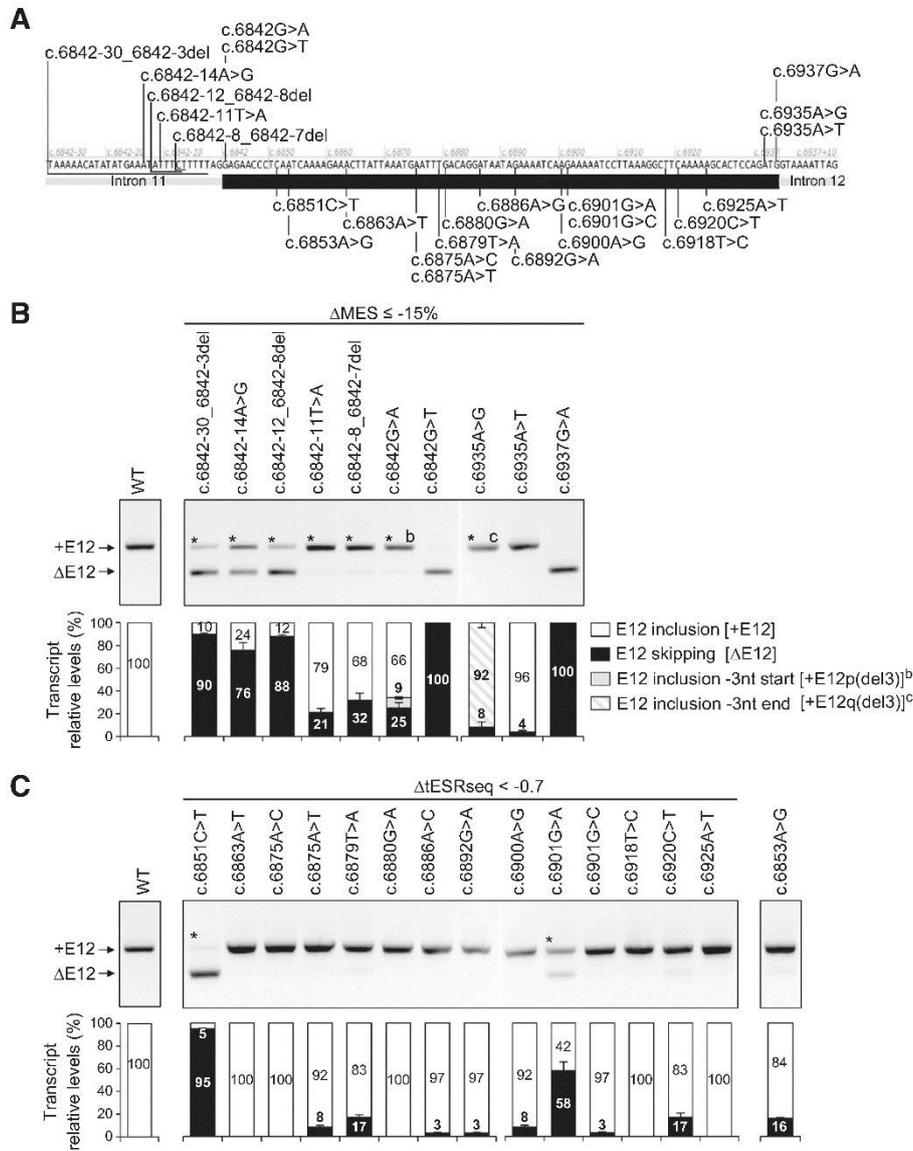


Figure 2. Identification in minigene assays of *BRCA2* exon 12 spliceogenic variants among SNVs predicted to affect splicing. **A**, Distribution of 24 variants predicted to affect *BRCA2* exon 12 splicing. The figure displays the nucleotide sequence of *BRCA2* exon 12 and part of its flanking intronic regions. The relative positions of the 10 selected variants predicted to affect *BRCA2* exon 12 splice sites ($\Delta\text{MES} \leq -15\%$) are indicated above the sequence, whereas the 14 selected variants that may impact *BRCA2* exon 12 splicing regulation based on ESR predictions ($\Delta\text{tESRseq} < -0.7$) are shown under the sequence. The position of the control spliceogenic variant, c.6853A>G, is also indicated. **B**, Impact on *BRCA2* exon 12 splicing of 10 variants predicted to alter splice sites ($\Delta\text{MES} \leq -15\%$). WT and mutant pCASP2-*BRCA2*-E12 minigene constructs were transiently expressed in HeLa cells. The splicing patterns of the RNA produced from these different minigenes were analyzed by RT-PCR, as described in Materials and Methods. The image shows the agarose gel electrophoresis of the RT-PCR products. The identities of the two major RT-PCR products, with (+E12) or without (Δ E12) exon 12 are indicated on the left of the gel. The other isoforms identified, b and c, correspond to the inclusion of exon 12 deleted either for the first [+E12p(del3)] or the last [+E12q(del3)] three nucleotides. Here, it is important to note that the gel images were not used for the quantification of the relative transcript levels because of the formation of heteroduplex between the different RNA isoforms (noted by the star symbol on the gel images), as revealed by band sequencing. Instead, the relative quantifications of the different RNA isoforms were determined by capillary electrophoresis of the fluorescence RT-PCR products (see Supplementary Fig. S5), as indicated in Materials and Methods. The data shown represent the average of at least three independent experiments. Error bars, SD values. **C**, Impact on *BRCA2* exon 12 splicing of 14 variants predicted to modify regulatory elements ($\Delta\text{tESRseq} < -0.7$). The effects on splicing of the variants selected on the basis of predictions of ESR alterations were assessed using the minigene reporter assay, as described in **B**. In parallel, we analyzed the previously described spliceogenic c.6853A>G variant (13). The image shows agarose gel electrophoresis of the RT-PCR products. For each minigene, the relative quantifications of the different RNA isoforms were determined as described in **B**. The data represent the average of at least three independent experiments. Error bars, SD values. nt, nucleotide.

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

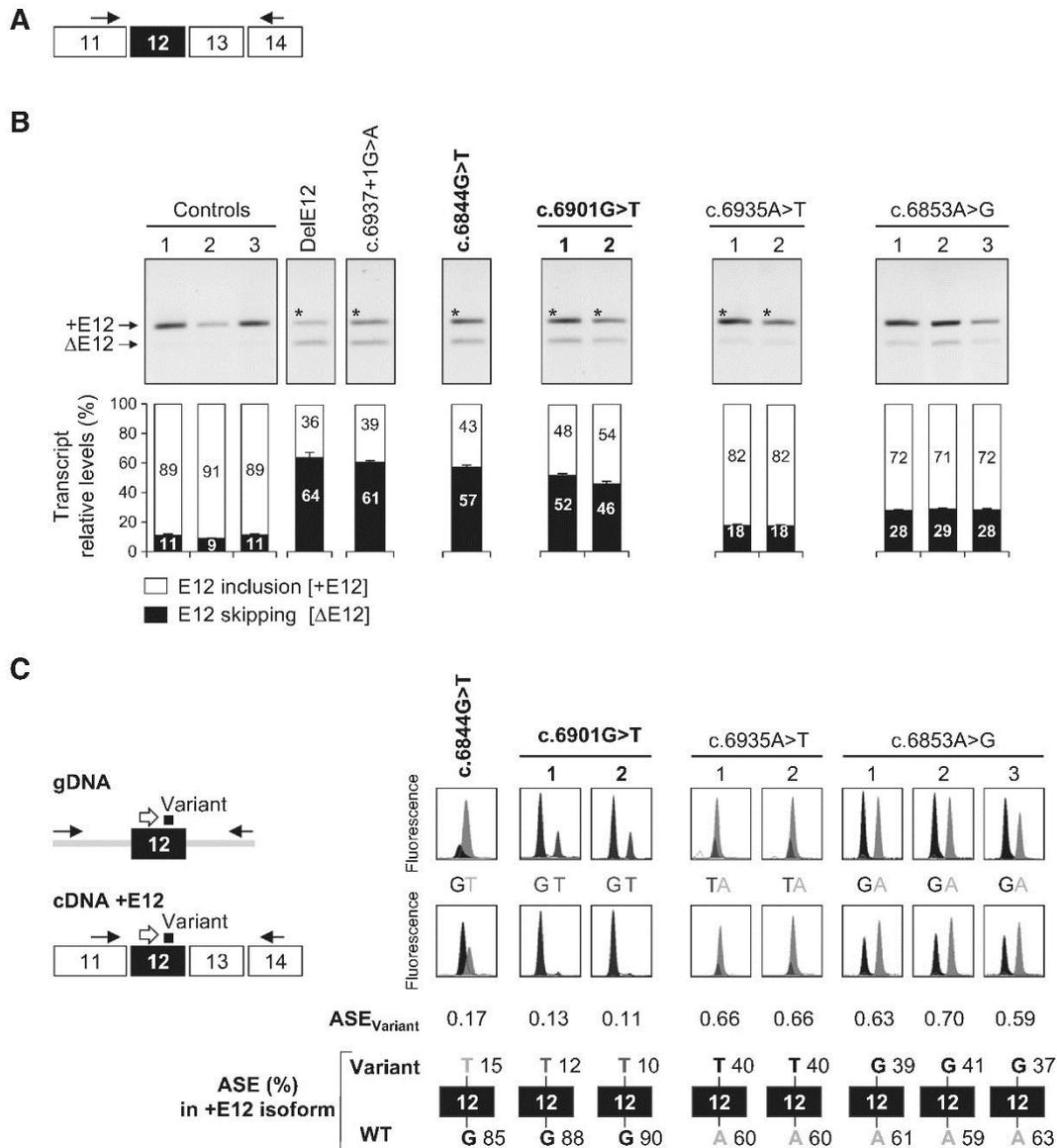


Figure 3. *BRCA2* exon 12 splicing patterns and ASE in control and patient LCLs. **A**, Location of the primers used for RT-PCR analysis of *BRCA2* exon 12 splicing patterns. Boxes indicate *BRCA2* exons 11-14 and arrows above indicate the position of primers used in RT-PCR analysis. **B**, RT-PCR analysis of *BRCA2* exon 12 splicing patterns. The *BRCA2* exon 12 splicing profiles were assessed by RT-PCR in LCLs of three healthy donors (controls 1-3) and of patients carrying selected variants. For the nonsense c.6844G>T and c.6901G>T (bold), the missense c.6935A>T and c.6853A>G variants, the analyses were conducted in LCLs of one, two, or three unrelated patients, as indicated. The image shows the agarose gel electrophoresis of the RT-PCR products. The identities of the two RT-PCR products, with (+E12) or without (ΔE12) exon 12, are indicated on the left of the gel. Here, it is important to note that the gel images were not used for the quantification of the relative transcript levels because of the formation of heteroduplex between the different RNA isoforms (noted by the star symbol on the gel images), as revealed by band sequencing. Instead, for each LCL, the relative quantifications of these RNA isoforms were determined by capillary electrophoresis of the fluorescence RT-PCR products (see Supplementary Fig. S6), as indicated in Materials and Methods. The data shown represent the average of at least three independent experiments. Error bars, SD values. **C**, ASE of transcripts including *BRCA2* exon 12 assessed by SNaPshot analysis. To measure ASE of transcripts with exon 12 (+E12) in patient LCLs, a SNaPshot quantitative primer extension assay was performed, as described in Materials and Methods. Genomic segment (gDNA) and cDNA containing *BRCA2* exon 12 (+E12) were amplified in parallel using specific primers (black arrows). The SNaPshot quantitative primer extension assay was then performed with a primer (white arrow) targeting the sequence immediately upstream the *BRCA2* exon 12 variant (■). Results from gDNA (top graphs) and cDNA (bottom graphs) are shown on the right. ASE measured using the variant (ASE_{Variant}) was then calculated after normalization of the cDNA peak area ratio (Variant/WT) to the corresponding gDNA peak area ratio. On the basis of these data, a schematic representation of the ASE, expressed as percentage of the variant and WT alleles, is also presented. Results are representative of three independent experiments.

Meulemans et al.

= 61%) was measured for a patient carrying c.6937+1G>A (Fig. 3B; Supplementary Fig. S6), congruent with this variant being responsible for complete E12 skipping. Remarkably, in LCL of a patient carrying the nonsense c.6844G>T variant, the relative levels of Δ E12 transcripts were greatly increased (Δ E12 = 57%; Fig. 3B; Supplementary Fig. S6). Similar results were obtained for LCLs of 2 patients carrying the nonsense c.6901G>T variant (Δ E12 = 52% and 46%, respectively; Fig. 3B; Supplementary Fig. S6). Treatment of these LCLs with puromycin did not significantly affect transcript proportions (Supplementary Fig. S7), suggesting that the increase of Δ E12 levels does not result from nonsense-mediated mRNA decay of the PTC-containing +E12 RNA. In contrast, in LCLs with c.6935A>T or c.6853A>G, a weak or moderate increase in exon skipping was observed (Fig. 3B; Supplementary Fig. S6).

Altogether, the results obtained in patient LCLs are in agreement with minigene data. Still, it is important to note that the WT minigene did not reproduce the alternative splicing of *BRCA2* E12 observed in control LCLs, most probably because of different cellular and/or genomic contexts. As a consequence, the minigene assay may underestimate the level of variant-induced exon skipping. It is also possible that some variants without an effect in the minigene assay exhibit a weak splicing impact in patient LCLs. Despite these limitations, the minigene assay represents a powerful approach for screening spliceogenic *BRCA2* E12 variants, as all the variant-induced effects detected by this method were confirmed in available patient LCLs.

ASE of *BRCA2* transcripts in patient LCLs

We took advantage of the quantitative SNaPshot assay to measure ASE in +E12 RNA by directly interrogating the presence of exonic variants. From the LCLs with the nonsense variants c.6844G>T or c.6901G>T, we observed an important allelic imbalance (AI) in +E12 transcripts in favor of the WT allele (Fig. 3C). In contrast, a moderate AI was detected in +E12 isoforms expressed in two LCLs with c.6935A>T and in three LCLs with c.6853A>G. Next, ASEs were measured in both +E12 and Δ E12 transcripts by using the c.7242A>G polymorphism located in *BRCA2* E14 (Fig. 4A–C). As expected, in control LCL, we observed, a nearly equal contribution of the two alleles in the expression of both RNA isoforms. In contrast, in LCL with c.6937+1G>A, a complete imbalance was observed in +E12 transcripts (Fig. 4B). Congruently, a strong AI was detected for Δ E12 transcripts, confirming that this isoform is predominantly produced from the minor allele, which, by deduction, also carries c.6937+1G>A (Fig. 4C).

In LCLs of 2 patients heterozygous for both the c.6901G>T nonsense variant and the E14 SNP, we confirmed that +E12 transcripts are predominantly produced from the major/WT allele (Fig. 4B), whereas Δ E12 transcripts are predominantly produced from the minor/variant allele (Fig. 4C). These results strongly suggest that the c.6901G>T nonsense variant is responsible for major E12 skipping in patient samples, in good agreement with the minigene data.

Functional characterization of *BRCA2* variants in mESC

For determination of the functional consequences of variant-induced enhanced production of the Δ E12 isoform, we employed the mESC-based *BRCA2* functional assay (8, 19, 20). We tested one missense (c.6842G>T), one IVS±1/2 (c.6937+1G>A), and three nonsense (c.6844G>T, c.6859A>T, and c.6901G>T) variants selected on the basis of their splicing effects in minigene assays. First, we showed that all five selected variants were able to rescue the cell-lethal phenotype induced after Cre-mediated deletion of the conditional mouse *Brca2* allele, as well as the two controls DelE12 and c.6853A>G (Fig. 5).

As described previously (13), the WT *BRCA2* gene expressed in this system reproduced E12 alternative splicing (Δ E12 = 18%), whereas the DelE12 variant produced solely Δ E12 transcripts (Fig. 6A; Supplementary Fig. S8). The analysis of mESC expressing c.6842G>T and c.6937+1G>A revealed almost complete or full E12 skipping, respectively (Fig. 6A; Supplementary Fig. S8). Importantly, the impact on E12 splicing of the two nonsense variants, c.6844G>T and c.6901G>T, were confirmed in mESC as they caused nearly full E12 skipping (Fig. 6A; Supplementary Fig. S8). In contrast, the third nonsense variant tested, c.6859A>T, exhibited a slight increase in the relative levels of Δ E12 transcripts, as compared with WT (Fig. 6A; Supplementary Fig. S8). For the known spliceogenic c.6853A>G variant (13), a more extensive increase of E12 skipping levels was observed.

Western blot analysis confirmed the expression of *BRCA2* protein for all variants after Cre-mediated deletion of the conditional *mBrca2* allele (Fig. 6B; Supplementary Fig. S9). It is important to note that the full-length and Δ E12 isoforms cannot be distinguished by Western blotting due to the small difference in size. A truncated *BRCA2* protein at the expected size (~260 kDa) was only detected for the nonsense variant c.6859A>T (p.Arg2287*) that expressed substantial amount of PTC-containing +E12 transcript. In contrast, no truncated protein isoforms were detected for the two other nonsense variants, c.6844G>T (p.Glu2282*) and c.6901G>T (p.Glu2301*), suggesting that these nonsense variants only produce *BRCA2* Δ E12 protein, in good agreement with the RNA data (Supplementary Table S3).

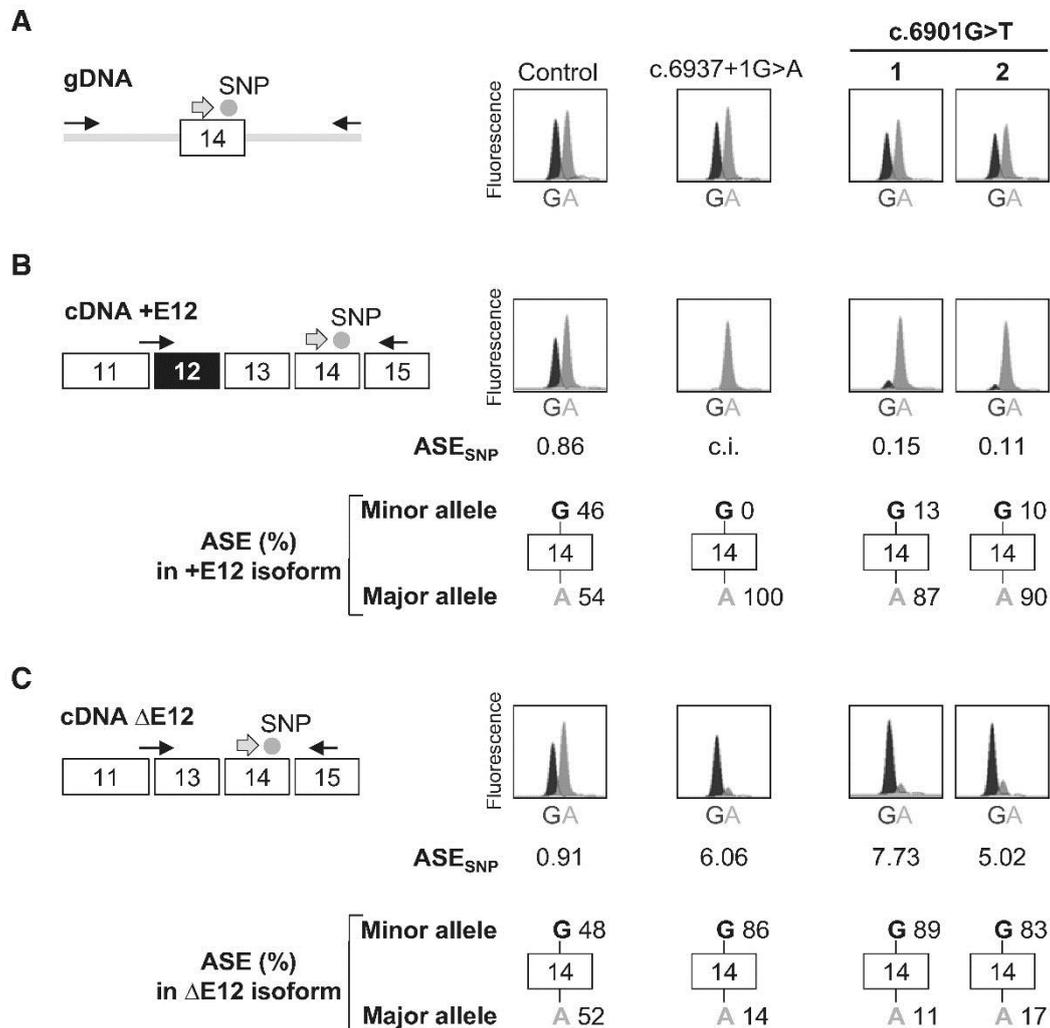
For functional characterization of the variants, we assessed their impact on HDR activity, the most prominent *BRCA2* tumor suppressor function (Fig. 6C). In a previous validation study, HDR ranges were defined to allow classification of variants in clinically relevant categories. *BRCA2* variants with >50% HDR activity are considered to be (likely) benign, while variants with <30% HDR activity are (likely) pathogenic (20).

The DelE12 variant displayed a 50% reduction in HDR activity as compared with WT. A similar reduction in HDR activity was observed for the previously described spliceogenic c.6853A>G missense variant (13), as well as for c.6842G>T, c.6937+1G>A, and the two nonsense variants, c.6844G>T and c.6901G>T. In contrast, the c.6859A>T nonsense variant displayed a more severe functional impact with a reduction of 70% in HDR activity and is the only variant in this set that is associated with a statistically significant reduction in HDR activity as compared with the DelE12 variant (Tukey multiple comparisons test comparing DelE12 with c.6859A>T, $P < 0.01$). The extent of functional impairment is in-line with the observation that this nonsense variant displayed the lowest relative levels of the “functional” Δ E12 transcript.

Previous data from Li and colleagues (13) showed no difference in homologous recombination between ES cells expressing WT and the *BRCA2* DelE12 and c.6853A>G transgenes, whereas here, we report 40%–50% reduction in HDR activity. The observed difference may be explained by differences in experimental set up both at the cell model level (single clones vs. polyclonal population of cells) and the assay used to determine HDR activity (targeted vector integration vs. HDR on a DR-GFP reporter assay). It is possible that the type or range of recombinational events assessed in these assays is not 100% identical. The DR-GFP reporter is an extensively used, well-established, and validated method to determine homologous recombination activity (20, 23). In addition, both our study and the data presented by Li and colleagues (13) show that there is no difference in cell viability of DelE12 and c.6853A>G variants as compared with WT.

Attenuation of the HDR pathway sensitizes cells to cisplatin and PARP inhibitors (20). Congruent with their substantial residual

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

**Figure 4.**

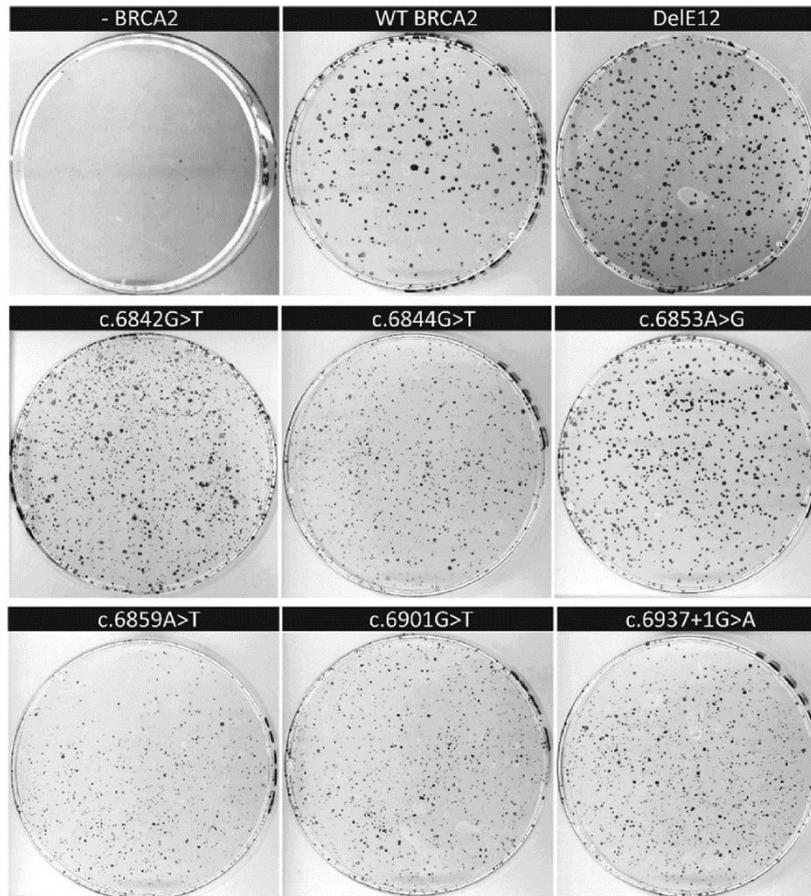
ASE of *BRCA2* transcripts with and without exon 12 in control and patient LCLs. **A**, SNaPshot results obtained from genomic DNA. As indicated in the left panel, the genomic segment encompassing *BRCA2* exon 14 was PCR amplified from gDNA of a control individual and patients carrying the heterozygous polymorphism located in *BRCA2* exon 14, c.7242A>G (SNP, •) by using specific primers (black arrows). The SNaPshot quantitative primer extension assay was then performed with a primer (gray arrow) targeting the sequence immediately upstream the SNP. Results shown on the right panel are representative of three independent experiments. **B**, ASE of transcripts containing *BRCA2* exon 12 (+E12). As shown in the left panel, +E12 cDNA was RT-PCR amplified from LCL RNA using specific primers (black arrows). The SNaPshot assay was next performed as described in **A**. Results are presented in the right panel. ASE measured in +E12 cDNA using the SNP (ASE_{SNP}) was calculated after normalization of the cDNA peak area ratio (minor allele/major allele) to the corresponding gDNA peak area ratio. Complete imbalance (c.i.) is indicated when only one allele was detected. A schematic representation of ASE_{SNP} , expressed as percentage of the minor and major alleles, is also presented. Results are representative of three independent experiments. **C**, ASE of transcripts lacking *BRCA2* exon 12 ($\Delta E12$). As indicated in the left panel, $\Delta E12$ cDNA was RT-PCR amplified from LCL RNA using specific primers (black arrows). The SNaPshot quantitative primer extension assay was performed as described in **A**. Results are presented in right panel. ASE measured in $\Delta E12$ cDNA using the SNP (ASE_{SNP}) was calculated after normalization of the cDNA peak area ratio (minor allele/major allele) to the corresponding gDNA peak area ratio. A schematic representation of ASE_{SNP} , expressed as percentage of the minor and major alleles, is also presented. Results are representative of three independent experiments.

capacity to perform HDR (i.e., $43\% \leq HDR \leq 61\%$), the major spliceogenic E12 variants displayed a relatively mild sensitivity to cisplatin and PARP inhibitor (talazoparib), whereas the c.6859A>T nonsense variant with a minor splicing effect and a more severe functional impact (i.e., $HDR = 26\%$) is associated with the highest drug sensitivity (Supplementary Fig. S10).

Clinical and family data of patients carrying E12 spliceogenic variants

We focused on families of 10 probands (P) that carry either variants inducing complete E12 skipping [DelE12 (P1), c.6842G>T (P2), and c.6937+1G>A (P3 and P4)] or nonsense variants that trigger major E12 skipping [c.6844G>T (P5 and P6) and c.6901G>T (P7–

Meulemans et al.

**Figure 5.**

Complementation phenotypes of *BRCA2* variants expressed in mESC. *Brca2*^{-/-}loxP; Pim1DR-GFP/WT cells without *BRCA2* (-*BRCA2*), or expressing *BRCA2* WT or variant were transfected with a Cre-GFP expression plasmid to induce loss of the conditional mouse *Brca2* allele. Upon Cre-recombinase expression, cells become *Brca2* deficient, which is lethal unless complemented by the expression of a (partially) functional *BRCA2* variant. Thirteen days post Cre-GFP transfection, culture dishes were stained with methylene blue.

P10); Fig. 7; Supplementary Table S4]. When taking into account women over 45 years of age, we observed a partial cosegregation of the variants with breast cancer. Indeed, besides probands, we identified: three carriers with breast cancer, five cancer-free carriers, and three noncarriers with breast cancer. Of note, cooccurrence of c.6842G>T with a pathogenic *BRCA2* exon 24 variant, c.9195_9196delinsAT (p.Phe3065_Gln3066delinsLeu*), was observed in P2. Yet, the chromosome phasing of these two variants remains unknown.

Discussion

It is now well established that mis-splicing triggered by specific SNVs is a common cause of Mendelian diseases (24). As a shift of paradigm, this study provides evidence for the first time in a cancer susceptibility gene, as far as we know, that variant-induced perturbation of splicing can rescue complete LoF of presumed pathogenic variants, at least partially.

For this demonstration, we used *BRCA2* E12 as a model system and took advantage of complementary experimental approaches to demonstrate that (i) a subset of presumed LoF IVS±1/2 variants as well as nonsense SNVs enhance in-frame skipping of *BRCA2* E12 and, as a consequence, (ii) lead to the production of an internally deleted but partially functional *BRCA2* protein. Altogether, these data call into question the clinical interpretation of specific spliceogenic hypomorphic IVS±1/2 and nonsense variants.

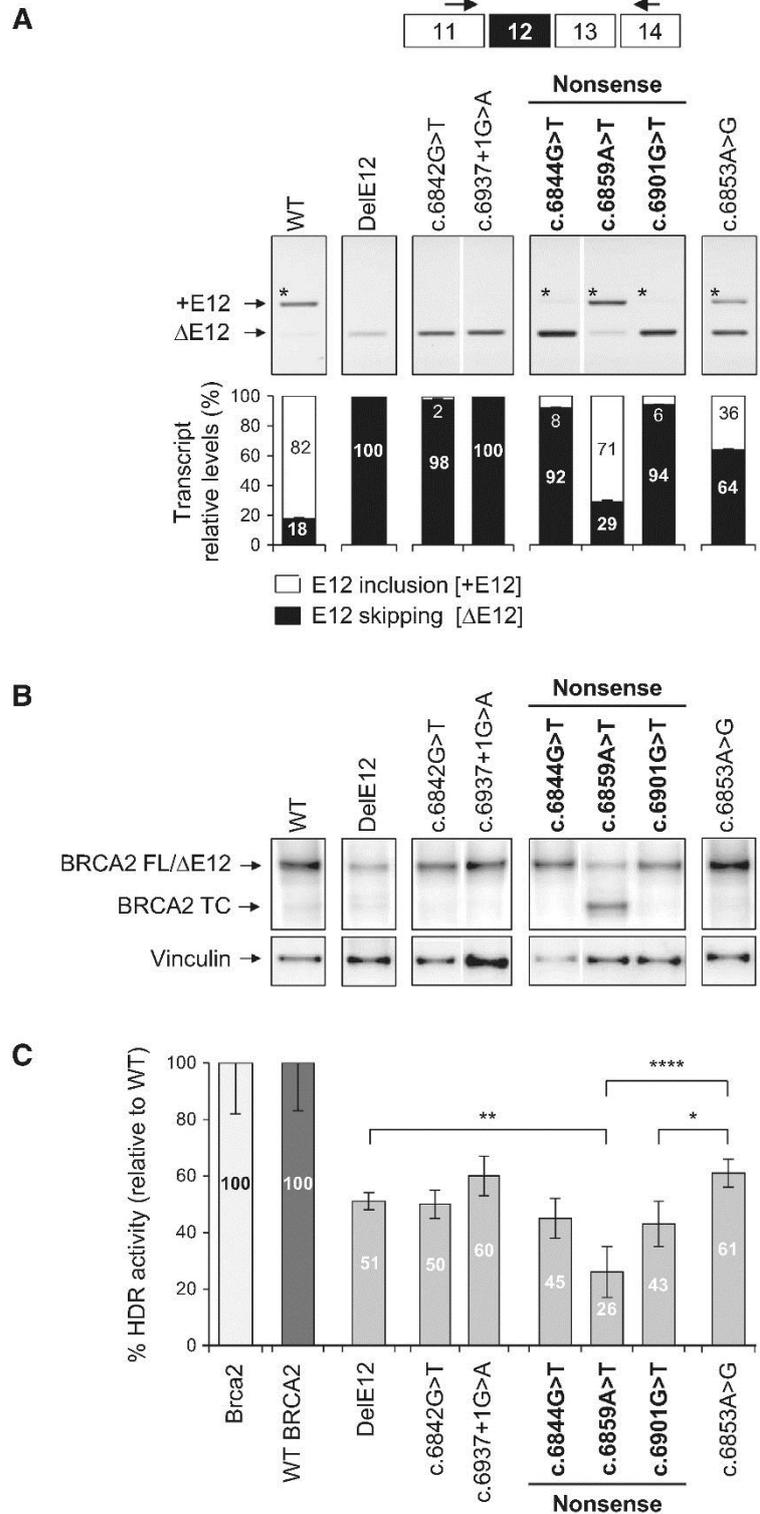
Whereas the severe impact of IVS±1/2 variants on *BRCA2* E12 splicing was expected given the disruption of the highly conserved 3' or 5' canonical ss, the effects of the five nonsense variants on splicing, albeit to different extents, were not anticipated and are most probably due to the modification of ESR, via either the destruction of exonic splicing enhancers (ESE) and/or the creation of exonic splicing silencers (ESS; refs. 11, 25). Interestingly, the two major spliceogenic nonsense variants described here, *BRCA2* c.6844G>T (p.Glu2282*) and c.6901G>T (p.Glu2301*), lie within purine-rich 5'-AGAA-3' motifs resembling those described as binding sites for the splicing activator protein Tra2-β1 (26). Most of the selected nonsense variants (4/5) promoted E12 skipping, albeit to different degrees, whereas none of the six frameshift variants tested had a splicing effect in minigene assays. Considering the relative limited number of variants in these two groups, the apparent specific association between nonsense mutations and E12 skipping might be coincidental as, in theory, small insertions or deletions within an exon could also fortuitously disrupt critical ESEs and/or create putative ESSs, both mechanisms potentially promoting exon skipping. Still, the association between nonsense mutations and E12 skipping could also reflect the fact that putative stop codon trinucleotides are enriched in ESS hexamers (27–29).

Nonsense-associated altered splicing has been previously described in cancer susceptibility genes (11). In particular, it has been shown that *BRCA1* c.5080G>T (p.Glu1694*) causes in-frame skipping of exon 18,

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

Figure 6.

Functional analysis by using a mESC-based assay of *BRCA2* variants that affect exon 12 splicing. **A**, *BRCA2* exon 12 splicing patterns. The location of the primers used for RT-PCR analysis of *BRCA2* exon 12 splicing patterns is presented on the top panel. Boxes indicate *BRCA2* exons 11-14 and arrows represent primers used in RT-PCR analysis. The *BRCA2* exon 12 splicing profiles were assessed by RT-PCR in mESC expressing either WT or mutant *BRCA2* from a genomic copy of the human gene. The identities of the two RT-PCR products, with (+E12) or without (Δ E12) exon 12, are indicated on the left of the gel. Here, it is important to note that the gel images were not used for the quantification of the relative transcript levels because of the formation of heteroduplex (noted by the star symbol on the gel images) between the different RNA isoforms, as revealed by band sequencing. Instead, the relative quantifications of different RNA isoforms were determined by capillary electrophoresis of the fluorescence RT-PCR products (see Supplementary Fig. S8). Data represent the average of at least three independent experiments. Error bars, SD values. **B**, *BRCA2* protein expression. Expression of *BRCA2* protein either full length (FL), or deleted for exon 12 (Δ E12) or truncated (TC) was detected by Western blot analysis. Vinculin was used as loading control. **C**, HDR activity. GFP signal was measured by flow cytometry in I-SceI-expressing cells 2 days post transfection. HDR activity is expressed as the percentage observed in *BRCA2* WT-expressing cells relative to *BRCA2* WT-expressing cells. *Brca2* represents the conditional *Brca2*^{-loxP/Pim1}^{DR-GFP/WT} cell line expressing endogenous mouse *Brca2*. Error bars indicate the SD of six independent GFP measurements per variant as represented by the dots. Tukey multiple comparisons test was applied to calculate *P* values comparing Δ E12 or c.6853A>G with exon 12 variants. *, *P* < 0.05; **, *P* < 0.01; ****, *P* < 0.0001.

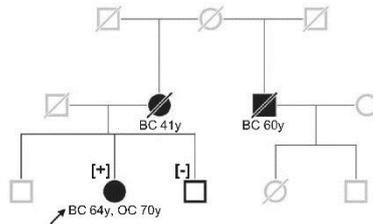


resulting in a nonfunctional protein, given the disruption of its first BRCT domain (30, 31). In contrast, for *BRCA2* E12 spliceogenic nonsense variants, the in-frame skipping of the PTC-containing exon

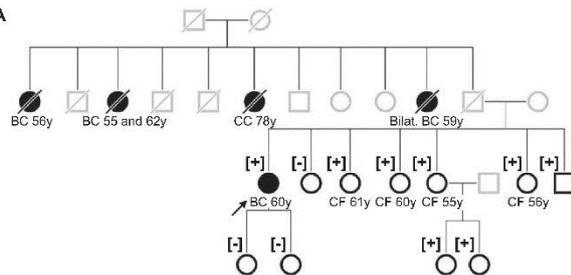
not only circumvents protein truncation, but also increases the production of an internally deleted partially functional isoform, as revealed in the mESC assay.

Meulemans et al.

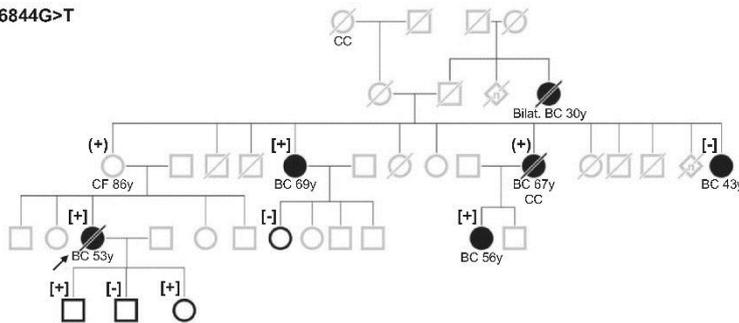
BRCA2 c.6842G>T
+ c.9195_9196delinsAT
(P2)



BRCA2 c.6937+1G>A
(P3)



BRCA2 c.6844G>T
(P5)



BRCA2 c.6901G>T
(P10)

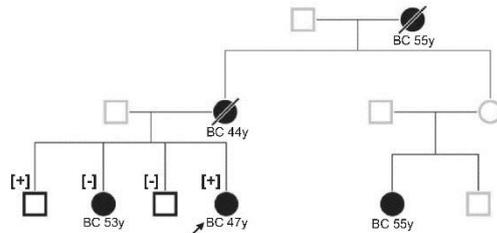


Figure 7.

Pedigrees of patients carrying *BRCA2* exon 12 spliceogenic variants. Here, the most informative pedigrees are presented. Proband referred to as P2, P3, P5, and P10, respectively, carriers of *BRCA2* c.6842G>T (p.Gly2281-Val), c.6937+1G>A, c.6844G>T (p.Glu2282*), and c.6901G>T (p.Glu2301*), are indicated by an arrow. Note that in patient P2, both *BRCA2* exon 12 variant (c.6842G>T) and *BRCA2* exon 24 pathogenic frameshift variant (c.9195_9196delinsAT) were identified. Filled symbols represent patients with cancer (BC, breast cancer; Bilat., bilateral; OC, ovary cancer; CC, colon cancer). Ages at cancer diagnosis are indicated. For some individuals, prophylactic salpingo-oophorectomy (PSO) was performed. Heterozygote carriers for the variant are indicated by the symbol [+], obligate carriers by the symbol (+), and noncarriers by the symbol [-]. CF, cancer free.

Recently, it has been shown that the mESC assay is a reliable method to predict pathogenicity of missense *BRCA2* variants (20). The complementation phenotype, the intermediate HDR activity, and the mild effect on cisplatin and PARP inhibitor sensitivity measured for the two major spliceogenic *BRCA2* E12 nonsense variants (c.6844G>T and c.6901G>T) suggest that these initially presumed null variants do not lead to complete LoF. The hypomorphic activity observed for these variants might confer moderate or low risks of breast cancer, as shown for previously characterized hypomorphic missense *BRCA1* and *BRCA2* variants (32–34). In contrast, the nonsense variant with the lowest level of E12 skipping (c.6859A>T) displays a significant

decrease in HDR activity comparable with a previously reported pathogenic missense variant (20). Consistent with the impact on HDR, this variant shows hypersensitivity to cisplatin-induced DNA damage and to PARP inhibition.

Besides the 15 presumed LoF variants, we extended the screening of *BRCA2* E12 spliceogenic variants to 25 additional SNVs (intronic, missense, and synonymous variants). Up to this date, only three missense variants have been reported to increase the level of alternative splicing of *BRCA2* E12 (13, 35, 36). This study not only confirmed those initial findings but, notably, uncovered 17 new spliceogenic variants affecting either 3'/5' ss or ESR.

BRCA2 Exon 12 Nonsense Spliceogenic Variants Retain Activity

In this study, we confirmed that the variant c.6853A>G promotes E12 skipping, as previously shown (13). However, this effect is partial: the mutant allele also significantly contributes to the production of the full-length +E12 transcript encoding the BRCA2 p.Ile2285Val protein. This observation does not call into question the nonpathogenic status of this variant. However, it prevents (i) from considering the neutral classification of this variant as an additional argument in favor of the functionality of the BRCA2 E12-deleted protein, as initially reported (13) and (ii) from extending this classification to all other BRCA2 E12 spliceogenic variants.

To apprehend the clinical relevance of BRCA2 E12 spliceogenic variants with respect to cancer risk, we analyzed patient clinical data and family history available within the French diagnostic network. Several observations (i.e., identification of both cancer-free carriers and noncarriers with breast cancer and cooccurrence with a pathogenic BRCA2 variant) could be in-line with the hypomorphic activity associated with these variants. Still, the interpretation of these spliceogenic variants cannot be inferred from these limited patient and family information and further larger-scale cosegregation and case-control studies will be needed to accurately evaluate cancer risks conferred by BRCA2 E12 spliceogenic variants.

To our knowledge, the existence of a rescue mechanism for nonsense variants through their direct impact on in-frame splicing has never been described for cancer susceptibility genes. Yet, it has been documented in genes involved in other Mendelian diseases. For instance, in the *DMD* gene, several truncating SNVs induce in-frame skipping of the PTC-containing exon, resulting in an internally deleted dystrophin, which exhibits partial function and consequently leads to a milder phenotype known as Becker muscular dystrophy (37–39). It is quite possible that presumed LoF BRCA2 E12 variants would confer severe, moderate, or minimal reduction in function, depending of their impact on splicing (i.e., levels of Δ E12 transcript), and would hence be associated with variable degrees of cancer susceptibility.

Of note, for certain SNVs, the BRCA2 Δ E12 RNA levels were somewhat variable in the three complementary approaches used in this study (i.e., minigene, patient LCL, and mESC assays), most probably because of different cellular and genomic contexts. Ideally, E12 splicing patterns should also be quantified in the patient relevant tissues (normal and tumoral breast tissues). Intriguingly, it has been previously reported that BRCA2 E12 alternative splicing is increased in some breast tumors as compared with their matched normal tissues (40) and upregulated in several mammary and ovarian cancer cell lines (41). Some of these cell lines could carry specific spliceogenic variants and provide attractive cellular model systems for future investigations on BRCA2 E12-deleted protein function. Importantly, during the preparation of this article, it has been shown that the highly conserved region encoded by E12 may be involved in the reproductive rather than the tumor-suppressing function of BRCA2 (42, 43). Indeed, the E12-encoded region (amino acid 2281–2313) is part of a binding domain for MEILB2/HSF2BP. This testis-expressed protein recruits BRCA2 to DNA double-strand breaks to ensure meiotic homologous recombination during spermatogenesis. Altogether, these data strongly suggest that BRCA2 E12 is essential to the successful completion of meiotic homologous recombination but not for the role of BRCA2 in meiotic homologous recombination, as a cancer suppressor.

In conclusion, this study provides for the first time experimental arguments to reevaluate the clinical interpretation of a subset of BRCA2 IVS±1/2 and nonsense spliceogenic variants that enhance in-frame skipping of a potentially nonessential exon. On the basis of these findings and following the ACMG (3) and ENIGMA (4) recommendations, we propose to reclassify these SNVs as variants of

unknown significance until further genetic analyses establish the cancer risks associated with these variants. This study emphasizes the requirement of complementary RNA and protein analyses to ensure comprehensive variant interpretation, with particular caution taken for presumed LoF variants located in in-frame dispensable exons. In the next-generation sequencing era, the main challenges for medical geneticists are not only the classification of the variants detected in patients but also the quantification of the associated risk, especially for variants retaining partial function. Indeed, our results highlight the need to move forward to a more accurate interpretation of variants in terms of both pathogenicity and penetrance.

Disclosure of Potential Conflicts of Interest

I. Tubeuf is a PhD student (paid consultant) at Interactive Biosoftware for the time period October 2015–September 2018 in the context of a public-private PhD project partnership between INSERM and Interactive Biosoftware (CIFRE fellowship #2015/0335). Astra Zeneca contributed financially to the COVAR study (<https://clinicaltrials.gov/ct2/show/NCT01689584>) coordinated by S. Caputo and D. Stoppa-Lyonnet. No potential conflicts of interest were disclosed by the other authors.

Consortia

Members of the French GGC network and of the group COVAR (Cosegregation of VARIants in the BRCA1/2 and PALB2 genes) are listed in the Supplementary Data.

Authors' Contributions

Conception and design: P. Gaildrat

Development of methodology: H. Vrieling, M.P.G. Vreeswijk

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Meulemans, R.L.S. Mesman, S.M. Caputo, S. Krieger, M. Guillaud-Bataille, N. Boutry-Kryza, J. Sokolowska, F. Révillon, C. Delnatte, H. Tubeuf, O. Soukarich, F. Bonnet-Dorion, M. Bronner, V. Bourdon, S. Lizard, P. Vilquin, M. Privat, C. Grout, F.M.G.R. Calléja, L. Golmard, D. Stoppa-Lyonnet, C. Houdayer, T. Frebourg, P. Gaildrat

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Meulemans, R.L.S. Mesman, S.M. Caputo, S. Krieger, H. Tubeuf, T. Frebourg, M.P.G. Vreeswijk, A. Martins, P. Gaildrat

Writing, review, and/or revision of the manuscript: L. Meulemans, R.L.S. Mesman, S.M. Caputo, L. Golmard, H. Vrieling, C. Houdayer, T. Frebourg, M.P.G. Vreeswijk, A. Martins, P. Gaildrat

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.M. Caputo, V. Caux-Moncoutier, A. Drouet, F.M.G.R. Calléja, C. Houdayer

Study supervision: H. Vrieling, M.P.G. Vreeswijk, A. Martins, P. Gaildrat

Acknowledgments

This work was financially supported by the Fédération Hospitalo-Universitaire (FHU) Normandy Centre for Genomic and Personalized Medicine (NGP) and the Gefuc. The Dutch Cancer Society KWF (U12012-5649) supported R.L.S. Mesman and F.M.G.R. Calléja. L. Meulemans was funded by a FHU-NGP PhD fellowship and H. Tubeuf by a CIFRE PhD fellowship (ANRT public-private partnership Inserm/Interactive Biosoftware#2015/0335). The COVAR study is supported by GENETI-CANCER association. The authors thank Françoise Charbonnier, Anne Rovelet-Lecrux, and Anne-Claire Richard (Inserm U1245, Rouen, France), Philippe Lafitte (Institut Curie, France), and Florence Rousselet (CHRU Nancy, France) for technical support; Jos Jonkers and Peter Bouwman (Cancer Institute, Amsterdam, the Netherlands) for the I-SceI-mCherry plasmid; Shyam Sharan (NCI, Frederick, MD) for the P12F7 conditional Brca2 knockout mES cell line; Maria Jasin (Memorial Sloan-Kettering Cancer Center, New York, NY) for the DR-GFP reporter plasmid, and Angela Rosania Solano (Faculty of Medicine, University of Buenos Aires, Buenos Aires, Argentina) for providing the patient samples.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 14, 2019; revised December 18, 2019; accepted February 6, 2020; published first February 11, 2020.

Meulemans et al.

References

- Mort M, Ivanov D, Cooper DN, Chuzhanova NA. A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat* 2008;29:1037–47.
- Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K-A, Mooij TM, Roos-Blom M-J, et al. Risks of breast, ovarian, and contralateral breast cancer for *BRCA1* and *BRCA2* mutation carriers. *JAMA* 2017;317:2402–16.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–23.
- Spurdle AB, Healey S, Devcreau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, et al. 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* 2012;33:2–7.
- de la Hoya M, Soukarieh O, López-Perollo I, Vega A, Walker LC, van Ierland Y, et al. Combined genetic and splicing analysis of *BRCA1* c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet* 2016;25:2256–68.
- Mazoyer S, Dunning AM, Scrova O, Dearden J, Puget N, Healey CS, et al. A polymorphic stop codon in *BRCA2*. *Nat Genet* 1996;14:253–4.
- Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian SV, et al. Functional evaluation and cancer risk assessment of *BRCA2* unclassified variants. *Cancer Res* 2005;65:417–26.
- Kuznetsov SG, Liu P, Sharan SK. Mouse embryonic stem cell–based functional assay to evaluate mutations in *BRCA2*. *Nat Med* 2008;14:875–81.
- Meeks HD, Song H, Michailidou K, Bolla MK, Dennis J, Wang Q, et al. *BRCA2* polymorphic stop codon K3326X and the risk of breast, prostate, and ovarian cancers. *J Natl Cancer Inst* 2016;108:djv315.
- Seo A, Steinberg-Shemer O, Unal S, Casadei S, Walsh T, Gumruk F, et al. Mechanism for survival of homozygous nonsense mutations in the tumor suppressor gene *BRCA1*. *Proc Natl Acad Sci U S A* 2018;115:5241–6.
- Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002;3:285–98.
- Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016;17:19–32.
- Li L, Biswas K, Habib LA, Kuznetsov SG, Hamel N, Kirchoff T, et al. Functional redundancy of exon 12 of *BRCA2* revealed by a comprehensive analysis of the c.6853A>G (p.I2285V) variant. *Hum Mutat* 2009;30:1543–50.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.
- Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, Russo JJ, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res* 2011;21:1360–74.
- Di Giacomo D, Gaildrat P, Abuli A, Abdal J, Frébourg T, Tosi M, et al. 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* 2013;34:1547–57.
- Soukarieh O, Gaildrat P, Hamich M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. *PLoS Genet* 2016;12:e1005756.
- Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* 2010;653:249–57.
- Hendriks G, Morolli B, Calléja FMGR, Plomp A, Mesman RLS, Meijers M, et al. An efficient pipeline for the generation and functional analysis of human *BRCA2* variants of uncertain significance. *Hum Mutat* 2014;35:1382–91.
- Mesman RLS, Calléja FMGR, Hendriks G, Morolli B, Misovic B, Devilee P, et al. The functional impact of variants of uncertain significance in *BRCA2*. *Genet Med* 2019;21:293–302.
- Packenthal JD, Yoshimatsu T, Zhang B, de Garibay GR, Colombo M, De Vecchi G, et al. Naturally occurring *BRCA2* alternative mRNA splicing events in clinically relevant samples. *J Med Genet* 2016;53:548–58.
- Davy G, Rousselin A, Goardon N, Castéra L, Harter V, Legros A, et al. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur J Hum Genet* 2017;25:1147–54.
- Guidugli L, Shimelis H, Masica DL, Pankratz VS, Lipton GB, Singh N, et al. Assessment of the clinical relevance of *BRCA2* missense variants by functional and computational approaches. *Am J Hum Genet* 2018;102:233–48.
- Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, et al. Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* 2017;49:848–55.
- Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 2008;14:802–13.
- Cléry A, Jayne S, Benderska N, Dominguez C, Stamm S, Allain FII-T. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2- β 1. *Nat Struct Mol Biol* 2011;18:443–50.
- Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. *Cell* 2004;119:831–45.
- Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res* 2011;21:1563–71.
- Ke S, Anquetil V, Zamalloa JR, Maity A, Yang A, Arias MA, et al. Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Res* 2018;28:11–24.
- Mazoyer S, Puget N, Perrin-Vidoz L, Lynch HT, Scrova-Similnikova OM, Lenoir GM. A *BRCA1* nonsense mutation causes exon skipping. *Am J Hum Genet* 1998;62:713–5.
- Liu H-X, Cartegni L, Zhang MQ, Krainer AR. A mechanism for exon skipping caused by nonsense or missense mutations in *BRCA1* and other genes. *Nat Genet* 2001;27:55–8.
- Spurdle AB, Whitley PJ, Thompson B, Feng B, Healey S, Brown MA, et al. *BRCA1* R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet* 2012;49:525–32.
- Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C, et al. *BRCA2* hypomorphic missense variants confer moderate risks of breast cancer. *Cancer Res* 2017;77:2789–99.
- Moghaddasi S, Meeks HD, Vreeswijk MP, Janssen LA, Borg Å, Ehrencrona H, et al. The *BRCA1* c.5096G>A p.Arg1699Gln (R1699Q) intermediate risk variant: breast and ovarian cancer risk estimation and recommendations for clinical management from the ENIGMA consortium. *J Med Genet* 2018;55:15–20.
- Brandão RD, van Roozendaal K, Tserpelis D, García EG, Blok MJ. Characterisation of unclassified variants in the *BRCA1/2* genes with a putative effect on splicing. *Breast Cancer Res Treat* 2011;129:971–82.
- Grodecká I, Lockerová P, Ravčuková B, Buratti E, Baralle FE, Dusček I, et al. Exon first nucleotide mutations in splicing: evaluation of *in silico* prediction tools. *PLoS One* 2014;9:e89570.
- Shiga N, Takeshima Y, Sakamoto H, Inoue K, Yokota Y, Yokoyama M, et al. Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J Clin Invest* 1997;100:2204–10.
- Ginjaar IB, Kneppers AL, v d Meulen J-DM, Anderson LV, Bremmer-Bout M, van Deutekom JC, et al. Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. *Eur J Hum Genet* 2000;8:793–6.
- Tuffery-Giraud S, Saquet C, Thorel D, Dissot A, Rivier F, Malcolm S, et al. Mutation spectrum leading to an attenuated phenotype in dystrophinopathies. *Eur J Hum Genet* 2005;13:1254–60.
- Bièche I, Lidereau R. Increased level of exon 12 alternatively spliced *BRCA2* transcripts in tumor breast tissue compared with normal tissue. *Cancer Res* 1999;59:2546–50.
- Rauh-Adelmann C, Lau K-M, Sabeti N, Long JP, Mok SC, Ito S-M. Altered expression of *BRCA1*, *BRCA2*, and a newly identified *BRCA2* exon 12 deletion variant in malignant human ovarian, prostate, and breast cancer cell lines. *Mol Carcinog* 2000;28:236–46.
- Zhang J, Fujiwara Y, Yamamoto S, Shibuya H. A meiosis-specific *BRCA2* binding protein recruits recombinases to DNA double-strand breaks to ensure homologous recombination. *Nat Commun* 2019;10:722.
- Brandma I, Sato K, van Rossum-Fikkert SE, van Vliet N, Sleddens E, Reuter M, et al. HSF2BP interacts with a conserved domain of *BRCA2* and is required for mouse spermatogenesis. *Cell Rep* 2019;27:3790–8.

Supplementary Data

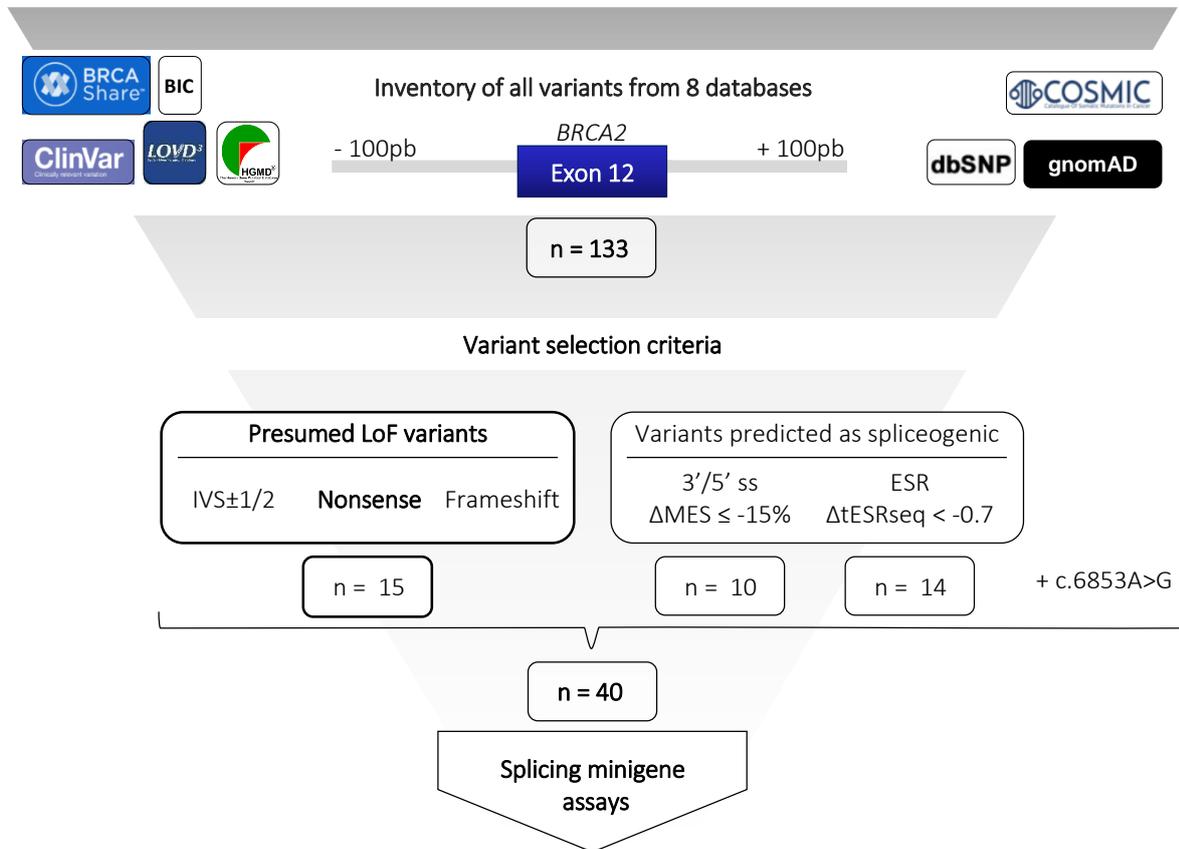


Figure S1. Variant selection from human variation databases

All variants located in *BRCA2* E12 and its 100 bp long flanking intronic sequences were extracted from the eight following human variation databases (last accessed: 05 September 2017): BRCA Share™, BIC (Breast Cancer Information Core), ClinVar, LOVD (Leiden Open Variation Database), HGMD (Human Gene Mutation Database), COSMIC (Catalogue of Somatic Mutations in Cancer), dbSNP (Single Nucleotide Polymorphism database) and gnomAD (genome Aggregation Database). A total of 133 variants were then collected. Firstly, all presumed LoF variants were selected, including intronic variants affecting the canonical intronic dinucleotides at 3'/5' ss (IVS±1/2) as well as nonsense and frameshift variants. In parallel, all other variants predicted to alter either 3'/5' ss or ESR, according respectively to the MaxEntScan algorithm ($\Delta\text{MES} \leq -15\%$) or the ESRseq scores ($\Delta\text{tESRseq} < -0.7$), were retained. In addition, c.6853A>G was included into the selection because it has been previously described as a *BRCA2* exon 12 spliceogenic variant (1). Altogether, this selection comprised 40 variants that were assessed for their impact on *BRCA2* exon 12 splicing in splicing minigene assays.

Figure S2

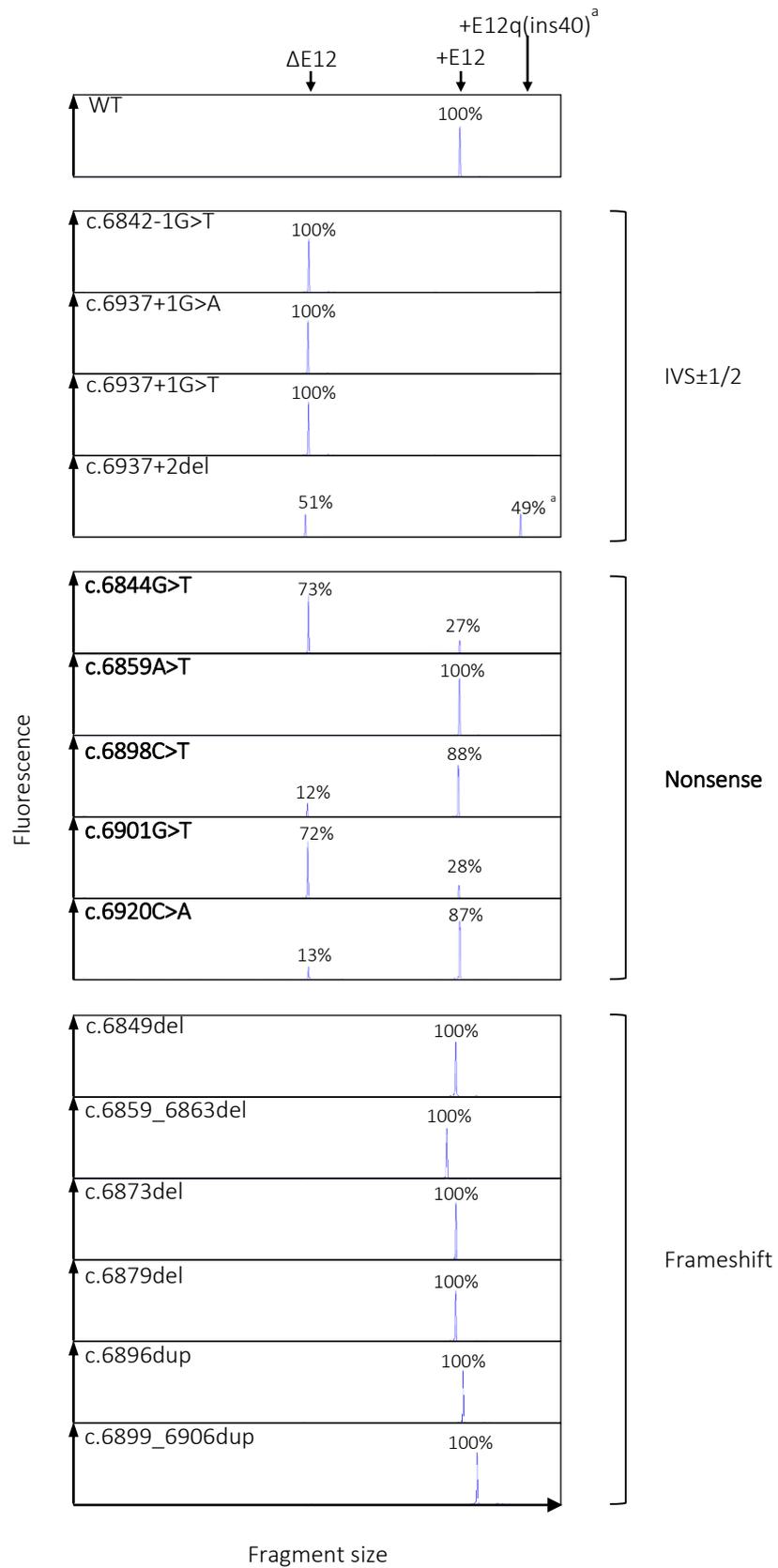


Figure S2. Capillary electrophoresis analyses of *BRCA2* exon 12 splicing patterns in minigene assays of presumed LoF variants

The *BRCA2* exon 12 splicing patterns of WT and mutant (IVS±1/2; nonsense; frameshift) minigenes expressed in HeLa cells were assessed by fluorescent RT-PCR, as described in Materials and Methods. In order to quantify the relative levels of each RNA isoform, the fluorescent RT-PCR products were resolved by capillary electrophoresis on an automated genetic analyzer. For each variant, one representative electropherogram is shown. The different fluorescence peaks correspond to transcripts either without ($\Delta E12$) or with (+E12) exon 12. An additional isoform (a) with the retention of the first 40 nucleotides of intron 12 (+E12q(ins40)) was detected in the context of c.6937+2del. Peak areas were used to quantify the relative levels of each transcript, expressed as percentage, and correspond to the average of at least three independent experiments (see Figure 1).

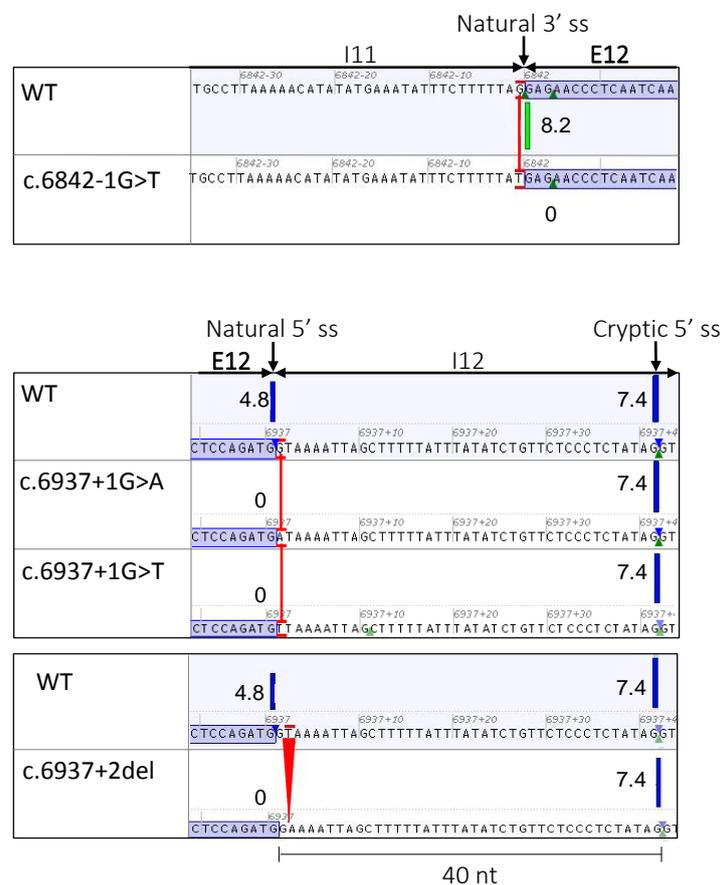


Figure S3. Bioinformatics predictions of 3'/5' ss alterations for variants located at position IVS±1/2 of *BRCA2* exon 12

Bioinformatics predictions of the variant effects on 3'/5' ss were based on the algorithm MaxEntScan (MES) (2), interrogated by using the integrated software Alamut Visual 2.9 (Interactive Biosoftware). The MES score of the 3' and 5' ss are indicated in the wild-type (WT) and variant contexts.

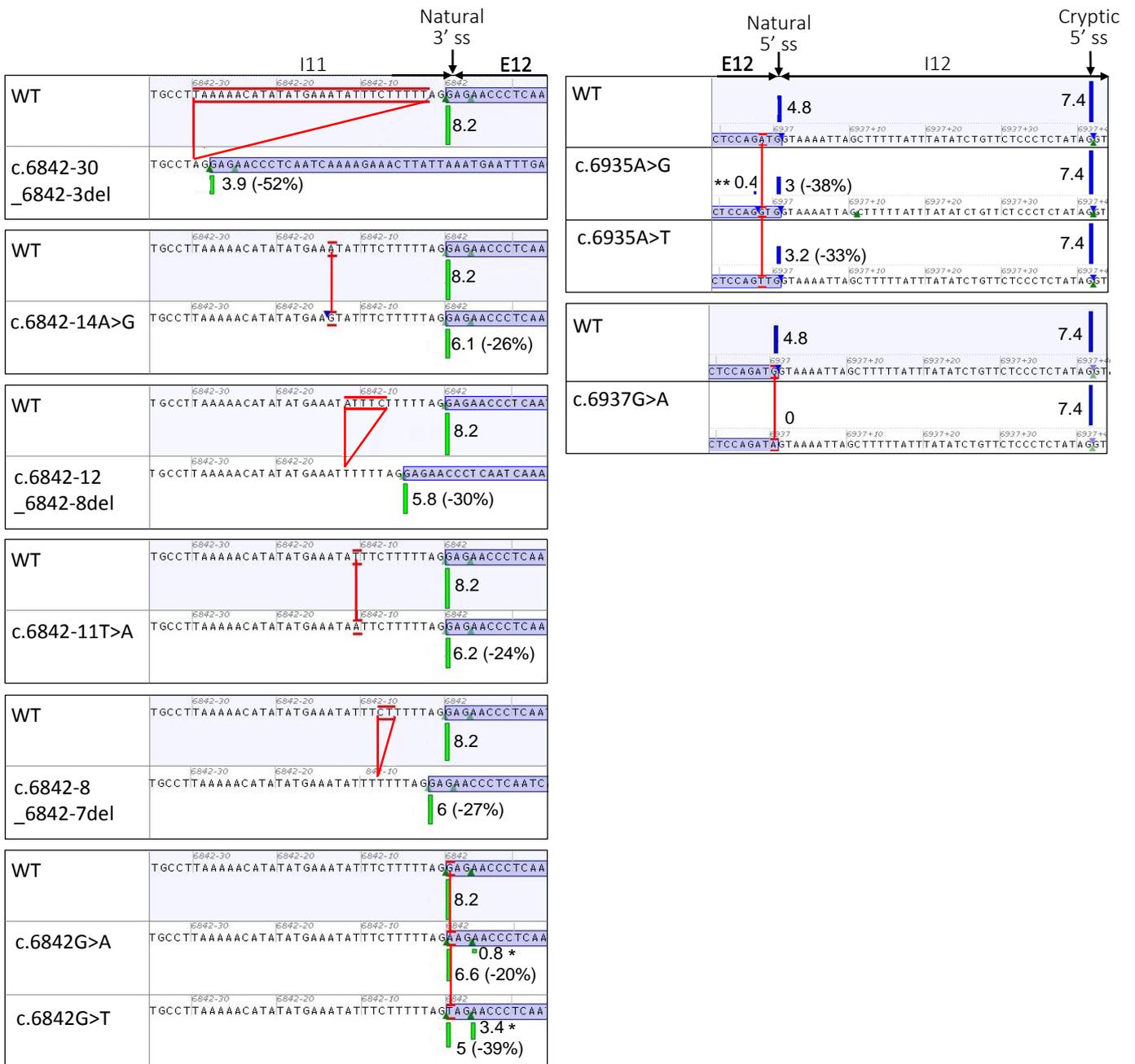


Figure S4. Bioinformatics analysis of variants predicted to alter 3'/5' ss of *BRCA2* exon 12 (Δ MES \leq -15%)

Bioinformatics predictions of the variant effects on 3' or 5' ss were performed based on the algorithm MaxEntScan (MES) (2), interrogated by using the integrated software Alamut Visual 2.9 (Interactive Biosoftware). For each variant, the MES score differences (Δ MES) between wild-type (WT) and variant contexts are indicated between brackets for 3'/5' natural ss. All these variants were selected for experimental analysis based on a Δ MES value inferior to -15%. In addition, the MES scores of new predicted 3' (*) or 5' (**) ss are indicated.

Figure S5

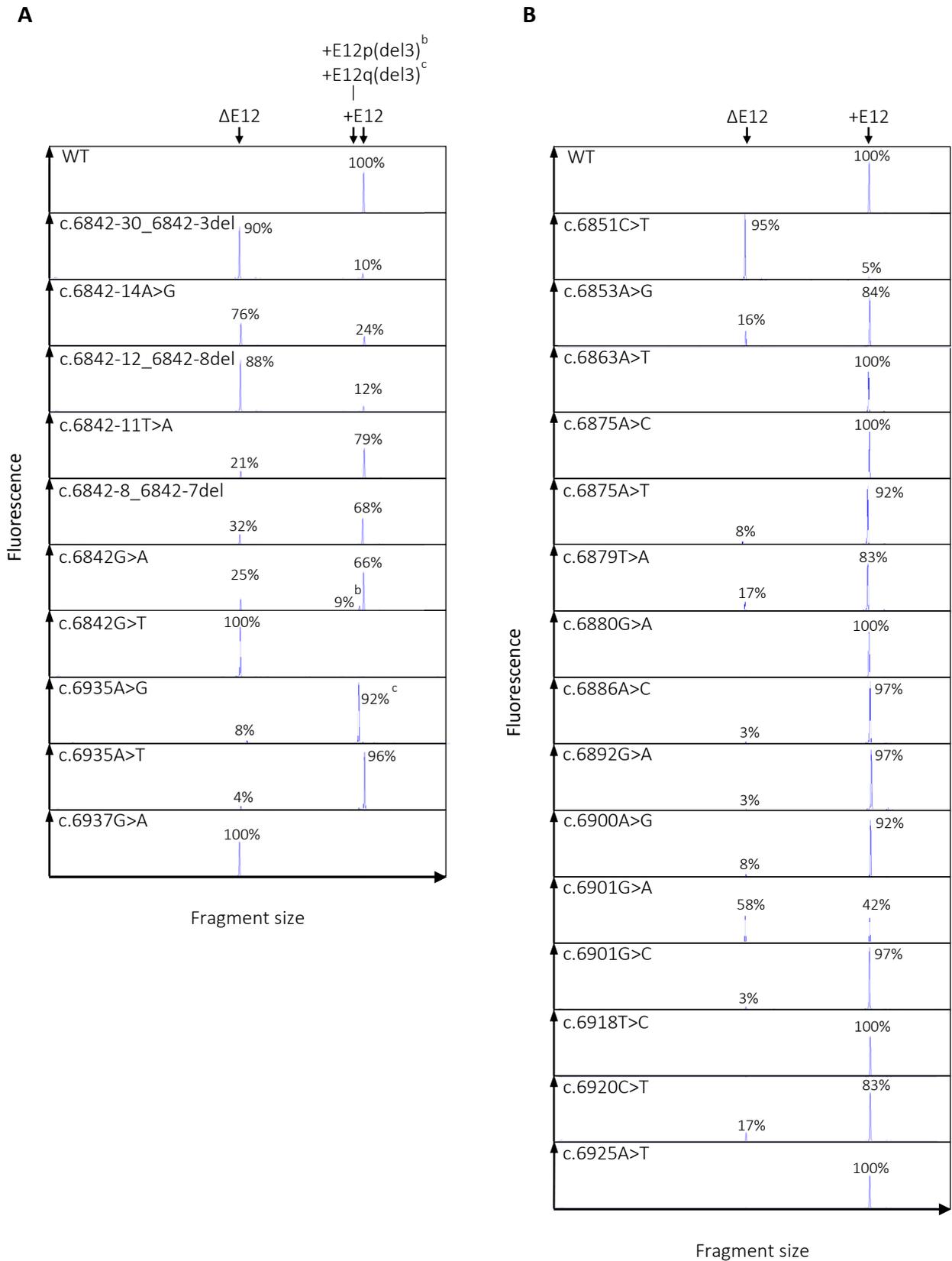


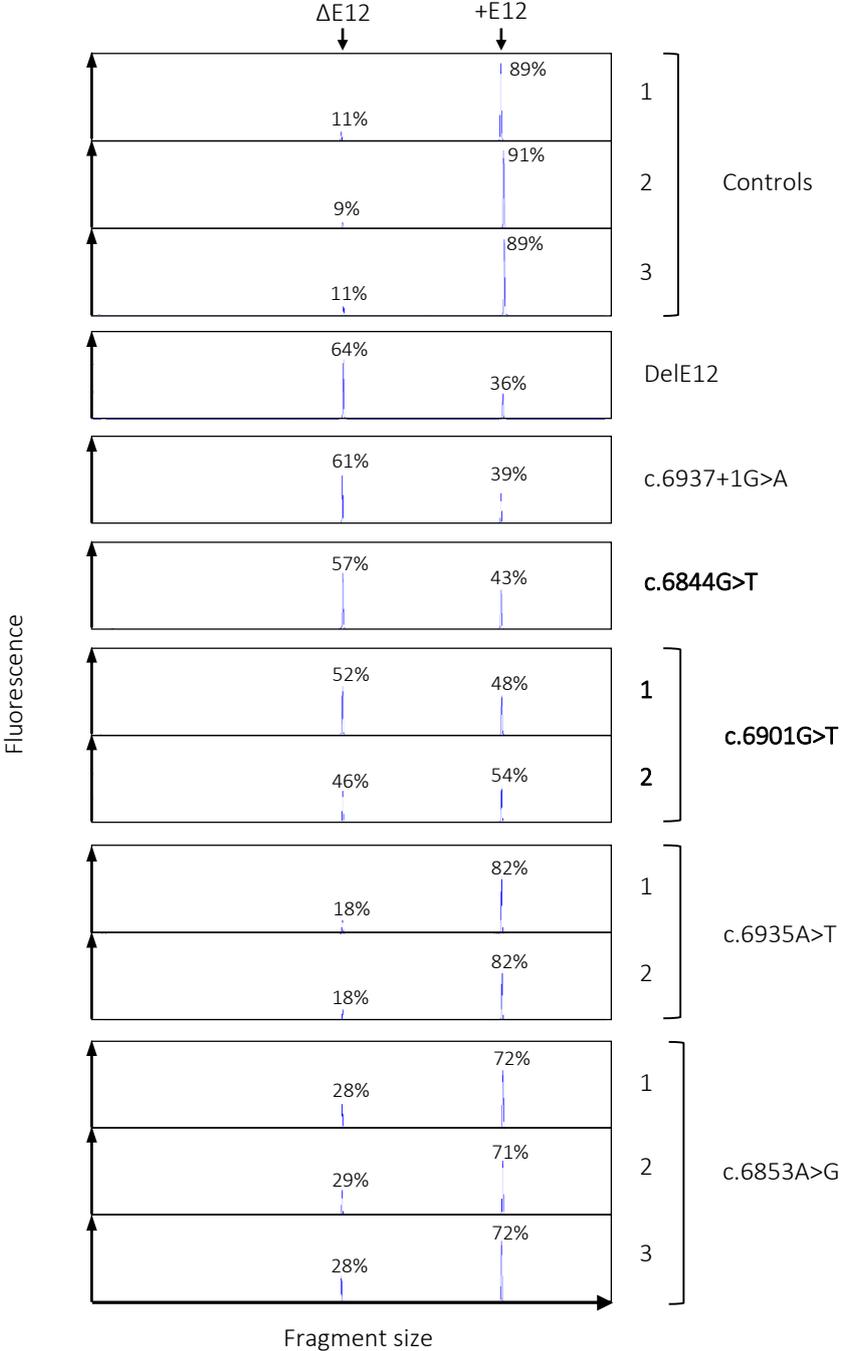
Figure S5. Capillary electrophoresis analyses of *BRCA2* exon 12 splicing patterns in minigene assays of variants predicted to alter 3'/5' ss (A) or ESR (B)

The *BRCA2* exon 12 splicing patterns of WT and mutant minigenes expressed in HeLa cells were assessed by fluorescent RT-PCR, as described in Materials and Methods. In order to quantify the relative levels of each RNA isoform, the fluorescent RT-PCR products were resolved by capillary electrophoresis on an automated genetic analyzer. For each variant, one representative electropherogram is shown. The different fluorescence peaks correspond to transcripts either without ($\Delta E12$) or with (+E12) exon 12. Additional isoforms, b and c, deleted either for the first (+E12p(del3)) or the last (+E12q(del3)) 3 nucleotides of exon 12 were detected for c.6842G>A and c.6935A>G, respectively. In parallel, the previously described spliceogenic c.6853A>G variant (1) was analyzed. Peak areas were used to quantify the relative levels of each transcript, expressed as percentage and correspond to the average of at least three independent experiments (see Figure 2).

Figure S6. Capillary electrophoresis analyses of *BRCA2* exon 12 splicing patterns in control and patient lymphoblastoid cell lines

The *BRCA2* exon 12 splicing patterns were assessed by RT-PCR in LCLs of 3 healthy donors (controls, 1 to 3) and of patients carrying selected variants. For the two nonsense, c.6844G>T and c.6901G>T (bold), the missense c.6935A>T and c.6853A>G variants, the analyses were conducted in LCLs of one, two or three unrelated carriers, as indicated. In order to quantify the relative levels of the different isoforms, the fluorescent RT-PCR products were resolved by capillary electrophoresis on an automated genetic analyzer. For each LCL, one representative electropherogram is shown. The different fluorescence peaks correspond to transcripts either without ($\Delta E12$) or with (+E12) exon 12. Peak areas were used to quantify the relative levels of each transcript, expressed as percentage, and correspond to the average of at least three independent experiments (see Figure 3).

Figure S6



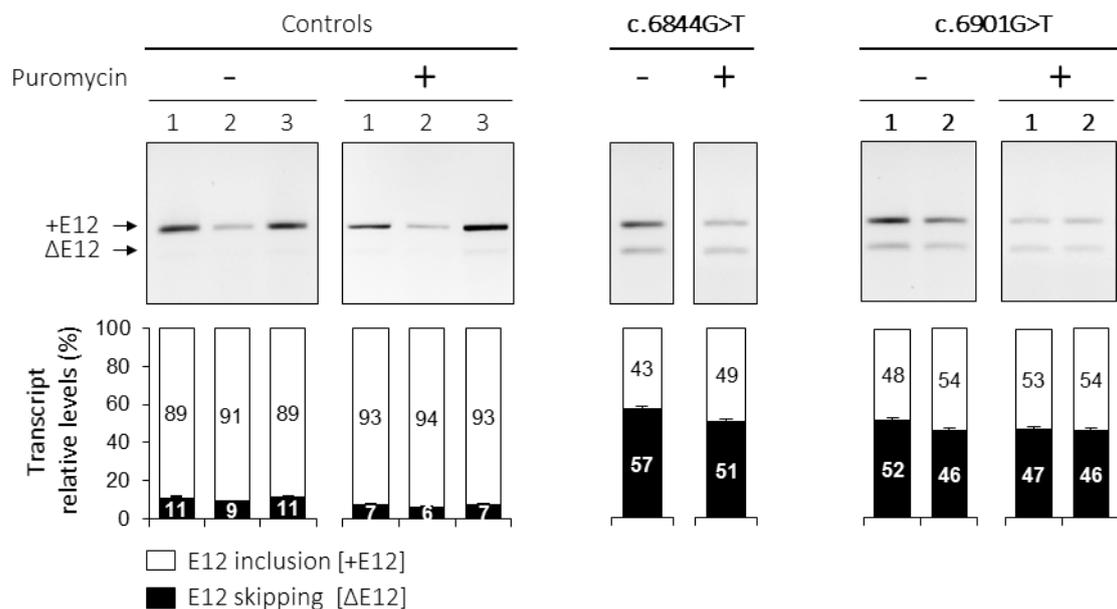


Figure S7. RT-PCR analysis of *BRCA2* exon 12 splicing patterns in puromycin- or mock-treated lymphoblastoid cell lines of control individuals and patients carrying the c.6844G>T or c.6901G>T nonsense variants

LCLs of three healthy donors (controls) and of a patient carrying the c.6844G>T nonsense variant as well as two unrelated patients carrying the c.6901G>T nonsense variant were cultured and treated (+) or mock-treated (-) with puromycin, as previously described (3). In these biological samples, the *BRCA2* exon 12 splicing patterns were assessed by RT-PCR, as described in Materials and Methods. The images show the agarose gel electrophoresis of the RT-PCR products. The identities of the two RT-PCR products, with (+E12) or without (ΔE12) exon 12, are indicated on the left of the gel. For each LCL, the relative abundances of these two RNA isoforms were determined by capillary electrophoresis of the fluorescent RT-PCR products, as indicated in Materials and Methods. The data shown under the images are expressed as percentage of transcript levels relative to all RT-PCR products and represent the average of at least three independent experiments. Error bars indicate standard deviation values.

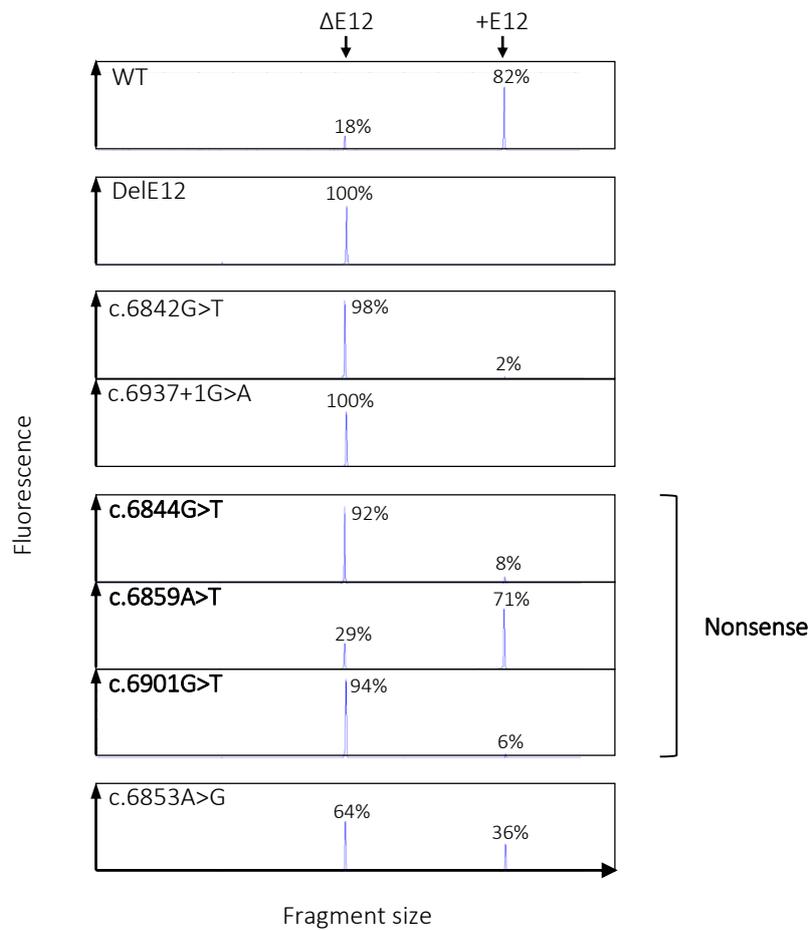


Figure S8. Capillary electrophoresis analyses of *BRCA2* exon 12 splicing patterns in variant-expressing mESC

The exon 12 splicing patterns of WT and mutant *hBRCA2* BAC constructs expressed in mESC were assessed by fluorescent RT-PCR, as described in Materials and Methods. In order to quantify the relative levels of each isoform, the fluorescent RT-PCR products were resolved by capillary electrophoresis on an automated genetic analyzer. For each variant-expressing mESC, one representative electropherogram is shown. The different fluorescence peaks correspond to transcripts either without ($\Delta E12$) or with (+E12) exon 12. Peak areas were used to quantify the relative levels of each transcript, expressed as percentage and corresponding to the average of at least three independent experiments.

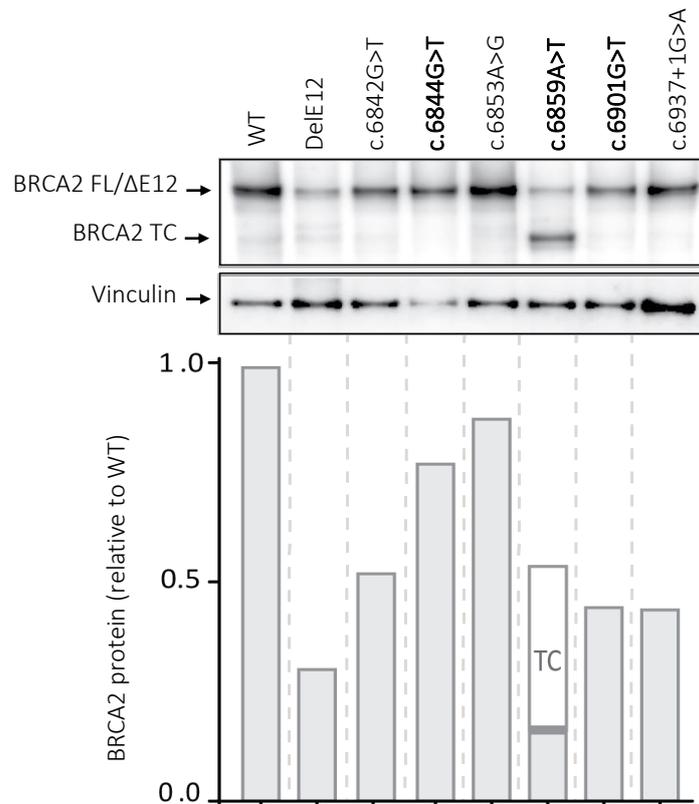


Figure S9. Quantitation of BRCA2 protein expression in mESC

Expression of BRCA2 full length (FL), without exon 12 (Δ E12) and truncated (TC) protein isoforms were detected by western blot analysis. Due to the small difference in size between the full length BRCA2 protein (BRCA2 FL protein isoform, 3418 aa) and the BRCA2 protein deleted for E12 (BRCA2 Δ E12 protein isoform, 3386 aa), these two protein isoforms are indistinguishable by western blot analysis. Vinculin was used as loading control. Chemiluminescent signals were captured by the Amersham Imager 600 (GE Healthcare Life Sciences) and band intensity quantification was performed on non-saturated images using ImageQuant TL software. The abundance of BRCA2 protein was normalized to the amount of vinculin in each lane (loading control). BRCA2 protein expression is presented relative to WT BRCA2 protein level.

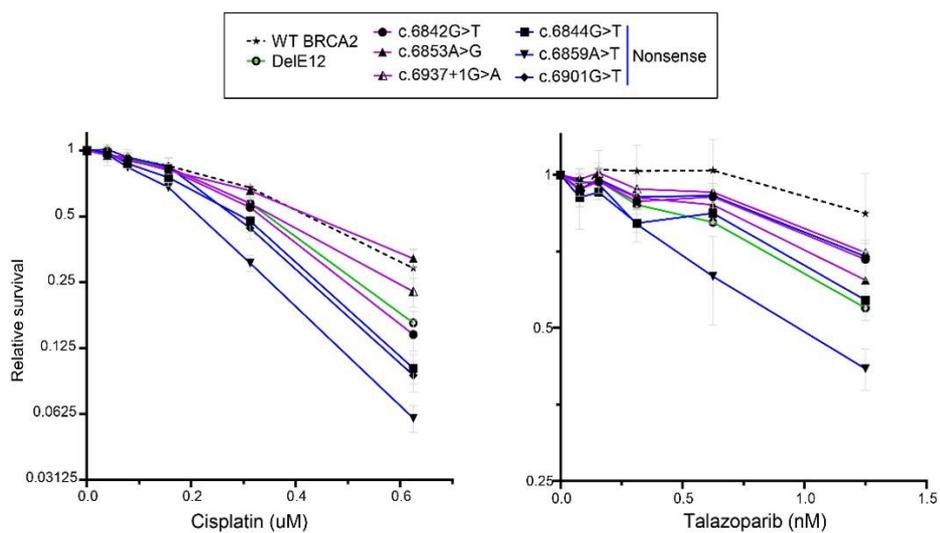


Figure S10. Sensitivity of BRCA2 variants to cisplatin and PARP inhibitors

Mouse ESC expressing *BRCA2* WT or variants were treated with cisplatin or PARP inhibitor (Talazoparib) and relative cell survival compared to untreated cells was determined as described in Materials and Methods. Data represent the average of four experiments per variant. Error bars indicate the standard error.

Supplemental references

1. Li L, Biswas K, Habib LA, Kuznetsov SG, Hamel N, Kirchhoff T, et al. Functional redundancy of exon 12 of *BRCA2* revealed by a comprehensive analysis of the c.6853A>G (p.I2285V) variant. *Hum Mutat* 2009;30:1543–50.
2. Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J Comput Biol* 2004;11:377–94.
3. Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. Aretz S, editor. *PLOS Genet* 2016;12:e1005756.

Supplemental Note

GGC Consortium

French HBOC diagnostic laboratories within the Group “Genetic and Cancer” (GGC, Unicancer)

Françoise Bonnet, Natalie Jones, Virginie Bubien, Michel Longy, Nicolas Sevenet: Institut Bergonie - Bordeaux; Sophie Krieger, Laurent Castera, Dominique Vaur: Centre François Baclesse - Caen; Maud Privat, Nancy Uhrhammer, Yves Jean Bignon: Centre Jean Perrin - Clermont-Ferrand; Sarab Lizard, Vincent Goussot: CHU de Dijon, Hôpital d'Enfants, Service de Génétique Médicale - Dijon; Aurélie Dumont, Françoise Revillion: Centre Oscar Lambret - Lille; Mélanie Léone, Nadia Boutry-Kryza, Olga Sinilnikova (deceased): Hospices Civils de Lyon and Centre Léon Bérard - Lyon; Audrey Remenieras, Violaine Bourdon, Tetsuro Noguchi, Hagay Sobol: Institut Paoli Calmettes – Marseille; Pierre-Olivier Harmand, Paul Vilquin, Pascal Pujol: Laboratoire de Biologie Cellulaire et Hormonale (CHU Arnaud de Villeneuve) - Montpellier; Philippe Jonveaux, Myriam Bronner, Joanna Sokolowska, Sarab Lizard: CHU de Nancy-Brabois - Vandoeuvre-les-Nancy; Capucine Delnatte, Virginie Guibert, Céline Garrec, Stéphane Bezieau: CHU - Institut de Biologie - Hôtel Dieu - Nantes; Erell Guillerm, Florence Coulet: Groupe hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Université Pierre et Marie Curie, Laboratoire d'Oncogénétique et Angiogénétique Moléculaire - Paris; Virginie Caux-Moncoutier, Lisa Golmard, Claude Houdayer, Dominique Stoppa-Lyonnet: Institut Curie - Paris; Fanny Barotel: Institut Jean Godinot - Reims; Danièle Muller: Centre Paul Strauss - Strasbourg; Christine Toulas: Institut Claudius Regaud - Toulouse; Etienne Rouleau, Marine Guillaud-Bataille, Brigitte Bressac-De Paillerets: Institut Gustave Roussy - Villejuif.

COVAR Group (COsegregation of VARIants in the *BRCA1/2* and *PALB2* genes)

(<https://clinicaltrials.gov/ct2/show/NCT01689584>)

Members	Center	City (France)
Florence Amram, Michèle Mathieu-Dramard	CHU Amiens	Amiens
Claire Le Hetet, Olivier Ingster	ICO - Centre Paul Papin Service de Génétique Médicale	Angers
Paul Gesta	Centre Hospitalier d'Angoulême Oncogénétique	Angoulême
Julie Mestre, Hélène Dreyfus	Institut Sainte-Catherine Oncogénétique	Avignon
Céline Populaire, Marie-Agnès Collonge-Rame, Jean-Luc Bresson,	CHU Besançon ; Hôpital Saint-Jacques ; Service de Génétique Biologique	Besançon
Michel Longy; Emmanuelle Barouk-Simonet, Virginie Bubien ; Henri Margot, Françoise Bonnet, Virginie Bubien, Natalie Jones, Nicolas Sévenet	Institut Bergonié	Bordeaux
Julie Tinat	CHU- Groupe Hospitalier Pellegrin	Bordeaux
Isabelle Mortemousque	CH Jacques Coeur	Bourges
Séverine Audebert-Bellanger	CHU de Brest	Brest
Pascaline Berthet; Agnès Hardouin, Sophie Krieger, Laurent Castéra, Dominique Vaur	Centre François Baclesse	Caen
Sandra Fert-Ferrer	Centre Hospitalier Hôtel Dieu Laboratoire de Génétique	Chambery
Nancy Uhrhammer, Maud Privat, Mathilde Gay- Bellile, Yves-Jean Bignon,	Centre Jean Perrin - Département d'Oncogénétique	Clermont-Ferrand
Vincent Goussot , Salima El Chehadeh ; Laurence Faivre-Olivier	CHU de Dijon Hôpital d'Enfants Service de Génétique Médicale	Dijon
Odile Béra	CHU de Fort de France Unité d'Oncogénétique	Fort de France, Martinique
Helene Dreyfus, Anne Béchet, Marie Bidard, Dominique Leroux	CHU de Grenoble	Grenoble
Hanitra Ranjatoelina	CHU Sud Réunion	La Réunion
Paul Gesta	Centre Hospitalier Saint-Louis Oncogénétique	La Rochelle
Valérie LAYET	Hôpital Flaubert Unité de Génétique	Le Havre
Claude Adenis, Françoise Revillion	Centre Oscar Lambret, Département de Cancérologie Générale	Lille
Sophie Lejeune, Sylvie Manouvrier-Hanu	CHRU LILLE Hôpital Jeanne de Flandre Service de Génétique Clinique	Lille
Laurence Vénat-Bouvet	CHU Dupuytren Oncogénétique	Limoges
Valérie Bonadona; Sylvie Mazoyer ; Olga Sininilkova, Mélanie Léone, Adrien Buisson, Nadia Boutry-Kryza, Christine Lasset	Centre Léon Bérard Lyon	Lyon
Sophie Giraud	Hospices civils de Lyon	Lyon
Hagay Sobol, Tetsuro Noguchi, Violaine Bourdon, Audrey Remenieras, François Eisinger, Catherine Noguès	Institut Paoli-Calmettes Département d'Oncologie Génétique	Marseille
Hélène Zattara	CHU La Timone, Département de Génétique	Marseille
Viorica Ciorna Monterrato, Julie Arel	CHR Metz	Metz
Jean-Marc Rey, Pierre-Olivier Harmand, Paul Vilquin, Virginie Galibert, Isabelle Coupier	CHU Arnaud de Villeneuve - Service de Génétique	Montpellier
Elisabeth Luporsi	Centre Alexis Vautrin; Oncogénétique	Nancy

Myriam Bronner, Joanna Sokolowska-Gillois, Christophe Philippe, Sarab Lizard, Philippe Jonveaux	CHU Nancy ; Hôpital Brabois ; Service de Génétique	Nancy
Capucine Delnatte, Virginie Guibert, Stéphane Bézieau, Estelle Cauchin, Caroline Abadie	ICO - Centre René Gauducheau Unité d'Oncogénétique	Nantes
Alain Lortholary, Alica Yvard	Centre Catherine de Sienne Oncogénétique	Nantes
Véronique Mari; Marc Frenay	Centre Antoine Lacassagne - Nice	Nice
Jean Chiesa, Audrey Combes	CHRU Caremeau, Oncogénétique	Nimes
Paul Gesta, Jeanne Triquard	Centre Hospitalier de Niort Oncogénétique	Niort
Isabelle Mortemousque	Hôpital de la Source Génétique	Orléans
Virginie Caux-Moncoutier, Lisa Golmard, Claude Houdayer, Muriel Belotti, Antoine Depauw, Caroline Leцерf, Marine Lementec, Anais Dupre, Bruno Buecher, Marion Gauthier-Villars, Sandrine Caputo, Dominique Stoppa-Lyonnet	Institut Curie site Paris Génétique Oncologique	Paris
Florence Coulet, Mathilde Warcoin, Erell Guillerm, Noemie Basset, Florent Soubrier	Groupe Hospitalier Pitié-Salpêtrière - Laboratoire d'Oncogénétique et d'Angiogénétique	Paris
Anne Fajac, Florent Soubrier	Hôpital Tenon Oncogénétique Clinique	Paris
Diane Molière, Pierre-Laurent Puig	HEGP	Paris
Marilyn Lackmy-Port-Lis	CHU de Pointe à Pitre	Pointe à Pitre, Guadeloupe
Paul Gesta, Brigitte Gilbert-Dussardier	CHU La Milétrie Oncogénétique	Poitiers
Nathalie Heil, Frédérique Carré-Pigeon	Polyclinique de Courlancy	Reims
Tan Dat Nguyen; Fanny Barotel, David Bonnerave; Aude-Marie Savoye	Institut Jean Godinot ; Service de Génétique et Biologie de la Reproduction	Reims
Dominique Gaillard ; Céline Poirsier ; Mozelle-Nivoix Monique	CHU de Reims ; Service de Génétique	Reims
Vanessa Colombert, Enora Lejeune, Caroline Abadie	Centre Eugène Marquis	Rennes
Annick Rossi, Isabelle Thevenet, Thierry Frebourg	CHU de Rouen Unité de Génétique Clinique	Rouen
Marine Lebrun, Fabienne Prieur	CHU Saint-Etienne Hôpital Nord	Saint-Etienne
Emmanuelle Fourme, Claire Saule, Ophélie Bertrand, Chrystelle Colas	Institut Curie site Hôpital René Huguenin Epidémiologie Clinique	Saint-Cloud
Jean-Pierre Fricker; Laurène Rebattu, Danièle Muller, Joseph Abecassis, Hélène Nehme-Schuster	Centre Paul Strauss ; Laboratoire de Biologie Tumorale	Strasbourg
Christine Maugard	Hopital Universitaire de Strasbourg	Strasbourg
Viviane Feillel; Norbert Lignon, Christine Toulas, Laurence Gladiëff	Institut Claudius Regaud Laboratoire d'Oncogénétique	Toulouse
Isabelle Mortemousque	CHU Bretonneau ; Service de Génétique ; Génétique	Tours
Monique Mozelle	CHG Troyes	Troyes
Marie-Noelle Bonnet-Dupeyron	Centre Hospitalier de Valence	Valence
Marine Guillaud-Bataille, Marine DImaria, Sophie Villebasse, Odile Cabaret, Alice Fievet, Etienne Rouleau, Brigitte Bressac-De Paillerets, Olivier Caron	Institut Gustave Roussy	Villejuif

Table S1. Description of primers used in this study

Primer's name ¹	Location	Sequence ²	Use	
B2I11BamHI-F	BRCA2	I11 GACCGATCTGTGAATTGACATTGAAGACTGACTTTAC	Genomic DNA PCR amplification for minigene cloning	
B2I12MluI-R		I12 GACCACGCCGTGAGTATAAAGAGGTCCTTGATTAGGCA		
B2E12 c.6842-30_6842-3del-F		I11/E12 ATTATTTGCCTAGGAGAACCC	Site directed mutagenesis pCAS2 : PCR1 ³	
B2E12 c.6842-30_6842-3del-R		I11/E12 GGGTTCTCCTAGGCAAATAAT		
B2E12 c.6842-14A>G-F		I11/E12 CATATATGAAGTATTTCTTTTAGGA		
B2E12 c.6842-12_6842-8del-F		I11/E12 ATATATGAAATTTTTAGGAGAACC		
B2E12 c.6842-12_6842-8del-R		I11/E12 GGTTCCTCAAAAAATTCATATAT		
B2E12 c.6842-11T>A-R		I11/E12 TCCTAAAAAGAAATTATTCATATATG		
B2E12 c.6842-8_6842-7del-F		I11/E12 TATGAAATATTTTTTAGGAGAACC		
B2E12 c.6842-1G>T-R		I11/E12 GTTCTCAAAAAAGAAATATTTTATA		
B2E12 c.6842G>A-F		I11/E12 ATTCTTTTAGAGAACCCTCAAT		
B2E12 c.6849del-R		I11/E12 ATTGAGGTTCTCCTAAAAAGAAAT		
B2E12 c.6851C>T-R		I11/E12 CTTTTGATTAGGGTTCTCTAA		
B2E12 c.6859_6863del-F		E12 CCCTCAATCAAACCTATTAATGA		
B2E12 c.6859_6863del-R		E12 TCATTTAATAAGTTGATTGAGGG		
B2E12 c.6863A>T-F		E12 AAAAGAACTCTTATTAATGAATTTGA		
B2E12 c.6873del-F		E12 AAACCTATTAAGAATTTGACAGG		
B2E12 c.6873del-R		E12 CCTGTCAAATCTTAATAAGTTT		
B2E12 c.6875A>T-F		E12 CTTATTAATGATTTTGGACAGG		
B2E12 c.6875A>C-R		E12 TCAAATGCATTTAATAAGTTTCTTTT		
B2E12 c.6879del-F		E12 TATTAATGAATTGACAGGATAATAG		
B2E12 c.6879T>A-F		E12 TTAATGAATTAGACAGGATAATA		
B2E12 c.6880G>A-R		E12 CTATTATCCTGTTAATTCATTTAATA		
B2E12 c.6886A>C-R		E12 TATTAGCCTGTCAAATTCATTTAA		
B2E12 c.6892G>A-F		E12 CAGGATAATAAAAAATCAAGA		
B2E12 c.6896dup-F		E12 GGATAATAGAAAATCAAGAAAAATC		
B2E12 c.6896dup-R		E12 GATTTTTCTTGATTTTCTATTATCC		
B2E12 c.6898C>T-F		E12 GAAAAATAGAAAAATCCTTAAAGGC		
B2E12 c.6899_6906dup-F		E12 TAGAAAATCAAGAAAAAAGAAAAATCCTTAAAG		
B2E12 c.6899_6906dup-R		E12 CTTAAGGATTTTCTTTTTTCTTGATTTTCTA		
B2E12 c.6900A>G-F		E12 ATAGAAAATCAGGAAAAATCCTTA		
B2E12 c.6900A>G-R		E12 TAAGGATTTTCTGATTTTCTAT		
B2E12 c.6901G>A-R		E12 GCCTTTAAGGATTTTTTTGATTTTC		
B2E12 c.6901G>C-R		E12 GAAAAATCAACAAAAATCCTTAAAGGC		
B2E12 c.6918T>C-F		E12 CCTTAAAGGCCTCAAAAAGCAC		
B2E12 c.6918T>C-R		E12 GTGCTTTTGAAGCCTTTAAGG		
B2E12 c.6920C>A-R		E12 CATCTGGAGTGCTTTTTTAAGCCTTAAAGG		
B2E12 c.6920C>T-F		E12 TAAAGGCTTTAAAAAGCACTC		
B2E12 c.6925A>T-R		E12 GAGTGCATTTTGAAGCCTTTA		
B2E12 c.6935A>G-F		E12/I12 CACTCCAGGTGGTAAAATTAGCT		
B2E12 c.6937G>A-F		E12/I12 CACTCCAGATAGTAAAATTAGC		
B2E12 c.6937+1G>T-R		E12/I12 AGCTAATTTTAAACATCTGGAGTG		
B2E12 c.6937+2del-R		E12/I12 GCTAATTTCCATCTGGAGTG		
pCASseq-F		pCAS2 minigene	JA GGGGTCAATAGCAGTGAGAG	Site directed mutagenesis pCAS2 (PCR1 and PCRII) and sequencing
pCASseq-R			JB GCTCCATTTACACAGGTAGAGA	
pCASKO1-F*	EA TGACGTGCGCCGCCATCAC		Analysis of BRCA2 E12 splicing patterns by RT-PCR and sequencing	
pCAS2-R	EB ATGGGTTGTTGAGTTGGTTGTC			

Minigene pCAS2-BRCA2 -E12

B2E11/12-F	BRCA2	E11/12	CCCTTATCTTAGTGG GAGAACCCCTCA	Analysis of BRCA2 E12 splicing patterns by RT-PCR and sequencing	LCL			
B2E11/13-F		E11/13	ATCTTAGTGG GCACAATAAAAG					
B2E15-R		E15	AAGAAGAAACAAGGCAACG					
B2i11BamHI-F		i11	GACCGGATCC TGTAATTGACATTGAAGACTGACTTTAC	Genomic DNA PCR amplification and sequencing				
B2i12Mlul-R		i12	GACCA CGCGI GAGTATAAAGAGGTCCTTGATTAGGCA					
B2i13-F		i13	AGCAAATGAGGGTCTGCAAC	Genomic DNA PCR amplification and sequencing				
B2i14-R		i14	GCAAAAATTCATCACACAAATTGTC					
B2E12-Snap-c.6844GT-R		E12	GTTTCTTTTGATTGAGGGTT	Analysis of allele-specific expression by SNaPshot				
B2E12-Snap-c.6853AG-R		E12	ATTTAATAAGTTCTTTTGA					
B2E12-Snap-c.6901GT-F		E12	ACAGGATAATAGAAAATCAA					
B2E12-Snap-c.6901GT-R		E12	TGAAGCCTTTAAGGATTTTT					
B2E12-Snap-c.6935AT-F		E12	GGCTTCAAAGCACTCCAG					
B2E14-Snap-c.7242AG-F		E14	CCACCTTTTAAACTAAATC					
B2E12-SDMc.6842G>T-F		BRCA2	i11/E12			GAAATATTTCTTTTAG I GAAACCCCTCAATCAAAG	Site directed mutagenesis pUC19	mESC
B2E12-SDMc.6842G>T-R	i11/E12		CTTTTGATTGAGGGTTCT A CTAAAAGAAATATTTTC					
B2E12-SDMc.6844G>T-F	i11/E12		GAAATATTTCTTTTAG G A I AACCCCTCAATCAAAGAAAC					
B2E12-SDMc.6844G>T-R	i11/E12		GTTTCTTTTGATTGAGGGTT A TCCTAAAAGAAATATTTTC					
B2E12-SDMc.6853A>G-F	i11/E12		GGAGAACCCCTC A GTCAAAGAACTTATTAATG					
B2E12-SDMc.6853A>G-R	i11/E12		CATTTAATAAGTTCTTTTG A CTGAGGGTCTCC					
B2E12-SDMc.6859A>T-F	E12		GAACCCCTCAATCAA A TGAAACTTATTAATG					
B2E12-SDMc.6859A>T-R	E12		CATTTAATAAGTTCT A TTTGATTGAGGGTTC					
B2E12-SDMc.6901G>T-F	E12		GGATAATAGAAAATCAA T AAAAATCCTTAAAGGC					
B2E12-SDMc.6901G>T-R	E12		GCCTTAAGGATTTT T ATTGATTTCTATTATCC					
B2E12-SDMc.6937+1G>A-F	E12/i12		AAGCACTCCAGATG A TAAATTAGCTTT					
B2E12-SDMc.6937+1G>A-R	E12/i12		AAAGCTAATTT A I I CATCTGGAGTGCTT					
B2i11 BamHI-F	i11		GACCGGATCC TGTAATTGACATTGAAGACTGACTTTAC	Transgene DNA PCR amplification and sequencing				
B2i12 Mlul-R	i12		GACCA CGCGI GAGTATAAAGAGGTCCTTGATTAGGCA					
B2E11-F*	BRCA2		E11	CCAAGTCATGCCACACATTC	Analysis of BRCA2 E12 splicing patterns	LCL and mESC		
B2E14-R			E14	ATTCTTGACCAGGTGCGGTA				

¹ F, forward; R, reverse; *, fluorescent primer

² Intronic and exonic sequences are indicated in grey and black, respectively.

The position of the variant is underlined. The double underlined sequences correspond to restriction sites for BamHI and Mlul.

³ The forward primers listed were used with the reverse primer pCAS-seq-R and the reverse primers listed were used with the forward primer pCAS-seq-F for mutagenesis.

Table S2. Variants selected in BRCA2 exon 12 and its flanking intronic regions

Selection criteria		Variant	Predicted amino acid change	Databases ²								
				BRCA-Share	BIC	ClinVar	LOVD	HGMD	COSMIC	dbSNP	gnomAD	
Presumed LoF variants n = 15	IVS±1/2 n = 4	c.6842-1G>T	p.?			4					4	
		c.6937+1G>A	p.?	5		5					5	√
		c.6937+1G>T	p.?			3						
		c.6937+2del	p.?					DM				
	Nonsense n = 5	c.6844G>T	p.Glu2282*	5								
		c.6859A>T	p.Arg2287*	5								
		c.6898C>T	p.Gln2300*	5								
		c.6901G>T	p.Glu2301*	5		5					5	
		c.6920C>A	p.Ser2307*	5								
	Frameshift n = 6	c.6849del	p.Ser2284Glnfs*7			4						√
		c.6859_6863del	p.Arg2287Leufs*4			5		DM			5	√
		c.6873del	p.Asn2291Lysfs*6					√				
		c.6879del	p.Phe2293Leufs*			5					5	
		c.6896dup	p.Asn2299Lysfs*41			5		DM			5	
		c.6899_6906dup	p.Ser2303Lysfs*5			5					5	√
Variants predicted as spliceogenic n = 24	In silico predictions of 3' or 5' ss alteration ΔMES ≤ -15% n = 10	c.6842-30_6842-3del	p.?		√	3					3	
		c.6842-14A>G	p.?		√	2					2	
		c.6842-12_6842-8del	p.?									√
		c.6842-11T>A	p.?		√	2					3	
		c.6842-8_6842-7del	p.?	3								
		c.6842G>A	p.Gly2281Glu								3	√
		c.6842G>T	p.Gly2281Val	3	√	3		DM?			3	
		c.6935A>G	p.Asp2312Gly									√
		c.6935A>T	p.Asp2312Val	1	√	c.i.	√	DM?			c.i.	√
		c.6937G>A	p.Gly2313Ser							√		
	In silico predictions of ESR alteration ΔTESRseq < -0.7 n = 14	c.6851C>T	p.Ser2284Leu							√		
		c.6863A>T	p.Asn2288Ile			3					3	
		c.6875A>C	p.Glu2292Ala								3	√
		c.6875A>T	p.Glu2292Val	3								
		c.6879T>A	p.Phe2293Leu					FP				
		c.6880G>A	p.Asp2294Asn	3								
		c.6886A>C	p.Ile2296Leu			3					3	√
		c.6892G>A	p.Glu2298Lys		√	3					3	
		c.6900A>G	p.= (p.Gln2300=)									√
		c.6901G>A	p.Glu2301Lys		√	3					3	
		c.6901G>C	p.Glu2301Gln			3					3	
		c.6918T>C	p.= (p.Ala2306=)			2					2	
		c.6920C>T	p.Ser2307Leu							√		
		c.6925A>T	p.Ser2309Cys							√		
Control ¹	c.6853A>G	p.Ile2285Val	2	√	1	√	FP			1	√	

¹ Variant previously described as spliceogenic

² Variant classification reported in BRCA-Share, ClinVar and dbSNP are indicated: 1, benign; 2, likely benign; 3, uncertain significance; 4, likely pathogenic; 5, pathogenic; c.i., conflicting interpretation = class 1, 2 or 3. For HGMD, the variant classification is indicated: FP, functional polymorphism; DM, disease causing mutation; DM?, disease causing mutation? In other databases: √, variant reported without classification.

Table S3. Overview of bioinformatics predictions and experimental data obtained for the 40 selected BRCA2 exon 12 variants

Selection criteria		Variant	Bioinformatics predictions	Splicing outcomes			mESC assay	
				LCL (biallelic) ²	Minigene assay (monoallelic)	mESC assay (monoallelic)	Protein expression ³	HDR activity
		Wild-type		ΔE12 = 11% ; 9% ; 11%	ΔE12 = 0%	ΔE12 = 18%	BRCA2 FL	100%
Presumed LoF variants n = 15	IVS±1/2 n = 4	c.6842-1G>T	ΔMES 3' ss = -100%		ΔE12 = 100%			
		c.6937+1G>A	ΔMES 5' ss = -100%	ΔE12 = 61%	ΔE12 = 100%	ΔE12 = 100%	BRCA2 ΔE12	60%
		c.6937+1G>T	ΔMES 5' ss = -100%		ΔE12 = 100%			
		c.6937+2del	ΔMES 5' ss = -100%		ΔE12 = 51% ; +E12q(ins40) = 49%			
	Nonsense n = 5	c.6844G>T	ΔtESRseq = -1.22	ΔE12 = 57%	ΔE12 = 73%	ΔE12 = 92%	BRCA2 ΔE12	45%
		c.6859A>T	ΔtESRseq = -1.20		ΔE12 = 0%	ΔE12 = 29%	BRCA2 TC (BRCA2 ΔE12)	26%
		c.6898C>T	ΔtESRseq = -2.38		ΔE12 = 12%			
		c.6901G>T	ΔtESRseq = -3.72	ΔE12 = 51% ; 46%	ΔE12 = 72%	ΔE12 = 94%	BRCA2 ΔE12	43%
		c.6920C>A	ΔtESRseq = -2.40		ΔE12 = 13%			
	Frameshift n = 6	c.6849del	ΔtESRseq = 0.56		ΔE12 = 0%			
		c.6859_6863del	ΔtESRseq = -2.20		ΔE12 = 0%			
		c.6873del	ΔtESRseq = -1.50		ΔE12 = 0%			
		c.6879del	ΔtESRseq = 0.23		ΔE12 = 0%			
		c.6896dup	ΔtESRseq = -0.59		ΔE12 = 0%			
		c.6899_6906dup	ΔtESRseq = -1.60		ΔE12 = 0%			
Variants predicted as spliceogenic n = 24	In silico predictions of 3' or 5' ss alteration ΔMES ≤ -15% n = 10	c.6842-30_6842-3del	ΔMES 3' ss = -52%		ΔE12 = 90%			
		c.6842-14A>G	ΔMES 3' ss = -26%		ΔE12 = 76%			
		c.6842-12_6842-8del	ΔMES 3' ss = -30%		ΔE12 = 88%			
		c.6842-11T>A	ΔMES 3' ss = -24%		ΔE12 = 21%			
		c.6842-8_6842-7del	ΔMES 3' ss = -27%		ΔE12 = 32%			
		c.6842G>A	ΔMES 3' ss = -20%		ΔE12 = 25% ; +E12p(del3) = 9%			
		c.6842G>T	ΔMES 3' ss = -39%		ΔE12 = 100%	ΔE12 = 98%	BRCA2 ΔE12	50%
		c.6935A>G	ΔMES 5' ss = -38%		ΔE12 = 8% ; +E12q(del3) = 92%			
		c.6935A>T	ΔMES 5' ss = -39%	ΔE12 = 18% ; 18%	ΔE12 = 4%			
		c.6937G>A	ΔMES 5' ss = -100%		ΔE12 = 100%			
	In silico predictions of ESR alteration ΔtESRseq < -0.70 n = 14	c.6851C>T	ΔtESRseq = -2.23		ΔE12 = 95%			
		c.6863A>T	ΔtESRseq = -0.82		ΔE12 = 0%			
		c.6875A>C	ΔtESRseq = -1.22		ΔE12 = 0%			
		c.6875A>T	ΔtESRseq = -1.32		ΔE12 = 8%			
		c.6879T>A	ΔtESRseq = -1.41		ΔE12 = 17%			
		c.6880G>A	ΔtESRseq = -1.01		ΔE12 = 0%			
		c.6886A>C	ΔtESRseq = -1.40		ΔE12 = 3%			
		c.6892G>A	ΔtESRseq = -0.98		ΔE12 = 3%			
		c.6900A>G	ΔtESRseq = -2.44		ΔE12 = 8%			
		c.6901G>A	ΔtESRseq = -2.44		ΔE12 = 58%			
		c.6901G>C	ΔtESRseq = -2.17		ΔE12 = 3%			
		c.6918T>C	ΔtESRseq = -0.74		ΔE12 = 0%			
		c.6920C>T	ΔtESRseq = -3.16		ΔE12 = 17%			
		c.6925A>T	ΔtESRseq = -1.06		ΔE12 = 0%			
Control ¹	c.6853A>G	ΔtESRseq = -0.30	ΔE12 = 28% ; 29% ; 28%	ΔE12 = 16%	ΔE12 = 64%	BRCA2 FL, BRCA2 ΔE12	61%	

¹ Variant previously described as spliceogenic

² When available, the data for unrelated controls or patients are indicated.

³ BRCA2 FL, full-length protein; BRCA2 Δ E12, protein deleted for E12; BRCA2 TC, truncated protein (p.Arg2287*). The minor protein isoform is indicated in parenthesis.

Table S4. Clinical and family data of patients carrying *BRCA2* exon 12 spliceogenic variants

Effect on splicing	Variant	Proband	Patient clinical and pathology data	Family data	Co-segregation data ¹			
					Carrier (Y) or non-carrier (N)	Breast cancer (BC), or other cancer (CC, PC) or cancer-free (CF)		
Complete skipping of <i>BRCA2</i> E12	c.6842-?_6937+7del (p.?)	P1	BC (67y) (SBR I, RE+, RP+, Her2-)	Sister 1	n.d.	BC (57y)		
				Sister 2	n.d.	BC (58y)		
				Maternal aunt 1	n.d.	BC (? y)		
				Maternal aunt 2	n.d.	BC (80y)		
	c.6842G>T (p.Gly2281Val) + c.9195_9196delinsAT (p.Phe3065_Gln3066delinsLeu*)	P2	BC (64y) (SBR I, RE+, RP-) OC (70y)	Mother	n.d.	BC (41y)		
				Half maternal uncle	n.d.	BC (60y)		
				Brother	N ²	CF (65y)		
				Paternal aunt 1	n.d.	BC (56y)		
	c.6937+1G>A (p.?)	P3	BC (60y) (SBR II, RE+, RP-, Her2+)	Paternal aunt 2	n.d.	BC (55y, 62y)		
				Paternal aunt 3	n.d.	Bilat. BC (59y)		
				Paternal aunt 4	n.d.	CC (78y)		
				Sister 1	N	CF (66y)		
				Sister 2	Y	CF (61y)		
				Sister 3	Y	CF (60y)		
				Sister 4	Y	CF (55y)		
				Sister 5	Y	CF (56y)		
Brother				Y	CF (51y)			
2 daughters				N	CF (41y, 39y)			
Niece 1, daughter of sister 4				Y	CF (38y)			
Niece 2, daughter of sister 4				Y	CF (33y)			
c.6844G>T (p.Glu2282*)	P5	BC (53y)	Sister	Y	CF (40y)			
			Daughter	Y	CF (30y)			
			Mother	Y ^o	CF (86y)			
			Maternal aunt 1	Y	BC (69y)			
c.6901G>T (p.Glu2301*)	P7	BC (58y) (RE+, RP+, Her2-)	Maternal aunt 2	Y ^o	BC (67y), CC			
			Maternal cousin (daughter of maternal aunt 2)	Y	BC (56y)			
Major increase of <i>BRCA2</i> ΔE12	c.6901G>T (p.Glu2301*)	P8	BC (61y)	Maternal aunt	N	BC (43y)		
				Paternal great aunt	n.d.	Bilat. BC (30y)		
				P6	BC (67y) (RE-, RP-, Her2-)	Sister	n.d.	BC (65y)
						Mother	n.d.	BC (50y)
						Sister	n.d.	BC (45y)
						Maternal cousin 1	n.d.	BC (77y)
				P9	CF (46y)	Maternal cousin 2	n.d.	BC (53y)
						Maternal grandmother	n.d.	BC (55-60y)
						Mother	n.d.	BC (56y, 64y)
						Maternal uncle 1	n.d.	PC
Maternal uncle 2	n.d.	PC						
Son	Y	CF (45y)						
P10	BC (47y) (SBR II, RE+, RP+)	Daughter	Y	CF (33y)				
		Maternal grandmother	n.d.	BC (70y)				
		Mother	n.d.	BC (35y)				
		Maternal aunt	n.d.	BC (31y)				
c.6901G>T (p.Glu2301*)	P10	BC (47y) (SBR II, RE+, RP+)	Maternal grandmother	n.d.	BC (55y)			
			Mother	n.d.	BC (44y)			
			Sister	N	BC (53y)			
			Brother	Y	CF (62y)			
c.6901G>T (p.Glu2301*)	P10	BC (47y) (SBR II, RE+, RP+)	Brother	N	CF (59y)			
			Maternal cousin	n.d.	BC (55y)			

Carrier (Y) with breast cancer (BC)
Carrier (Y) and Cancer-free (CF)
Non-carrier (N) with breast cancer (BC)

CF, cancer-free; BC, Breast Cancer; Bilat., Bilateral; OC, Ovary Cancer; CC, Colon Cancer; PC, Prostate Cancer; Y, carrier of the variant; Y^o, obligate carrier of the variant; N, non-carrier of the variant; N², negative for both variants; ¹Co-segregation data relative to women over 45 of age are indicated in bold and in colors as specified on the right.

II. Caractérisation fonctionnelle des variations à l'origine d'une anomalie en phase de l'épissage : le modèle *MSH2*

L'évaluation de l'impact des variations $IVS\pm 1/2$ et non-sens sur l'épissage de l'exon 12 de *BRCA2* nous a permis de valider notre hypothèse selon laquelle certaines variations supposées pathogènes, dans les gènes de prédisposition aux cancers, sont responsables du maintien d'une fonction du fait de leur impact sur l'épissage. Cette étude a donc apporté des arguments forts en faveur de l'existence d'un mécanisme de sauvetage de certaines variations supposées pathogènes, remettant ainsi en cause l'interprétation biologique et clinique de ces variations. Forts de ces résultats, nous avons entrepris d'étendre cette question à d'autres gènes de prédisposition aux cancers. Notre choix de modèle s'est alors porté sur *MSH2*, un des gènes majeurs de prédisposition au syndrome de Lynch, impliqué dans le système MMR (*MisMatch Repair*). A ce jour, dans ce gène, les variations à l'origine d'un défaut d'épissage en phase et leurs conséquences fonctionnelles n'ont pas fait l'objet d'investigations alors que ces modifications sont, en théorie, susceptibles de présenter un caractère neutre ou pathogène.

Notre stratégie a donc consisté à tester l'impact sur l'épissage de variations $IVS\pm 1/2$ localisées au niveau des exons en phase de *MSH2* ainsi que celles susceptibles d'induire d'autres types de modifications de l'épissage en phase. A l'aide d'un test fonctionnel indicateur d'anomalies de l'épissage basé sur l'utilisation de minigènes, nous avons caractérisé la nature et la sévérité de ces défauts d'épissage en phase et mis en évidence trois biotypes distincts : (i) des sauts d'exon, (ii) des délétions de portions exoniques et (iii) des rétentions introniques. Ces anomalies peuvent donc conduire à la production de protéines *MSH2* présentant des délétions ou insertions internes de taille plus ou moins importante dans différents domaines fonctionnels. Afin d'évaluer les conséquences de telles modifications sur l'activité de la protéine *MSH2*, nous avons mis à profit un essai fonctionnel basé sur la tolérance cellulaire aux lésions de l'ADN induites par des agents méthylants. Cette analyse a été menée en collaboration avec le Dr. Muleris et le Dr. Coulet, dans le cadre d'une mobilité de 2 mois (Service de génétique, Hôpital Universitaire Pitié-Salpêtrière, Paris). Dans cet essai, l'évaluation de la survie cellulaire en réponse à un agent méthylant a révélé que toutes les isoformes en phase étudiées ici sont associées à une réponse similaire à celle mise en évidence pour des variations pathogènes. Cette approche nous a donc permis de démontrer que toutes les variations à l'origine d'un défaut d'épissage en phase analysées dans le cadre de cette étude, conduisent à la perte de fonction de la protéine *MSH2*.

Ces résultats ont permis de montrer que, d'un point de vue fonctionnel, la protéine MSH2 ne tolère aucune des modifications insertion/délétion analysées ici, qu'elles soient de taille importante ou restreinte à quelques acides aminés, contrairement à ce qui a été démontré pour BRCA2. La perte de fonction observée peut être la conséquence de l'altération d'un domaine fonctionnel spécifique ou d'un défaut de conformation de la protéine qui aboutit à sa dégradation. Cette étude représente la première caractérisation fonctionnelle au niveau protéique des variations IVS±1/2 splicéogéniques dans un gène *MMR*. Jusqu'à maintenant, ce type d'analyse était principalement réservée aux variations de type faux-sens. En démontrant qu'aucune des isoformes insertion/délétion testées ne permet le maintien d'une fonction, nous confirmons le caractère pathogène de ces variations. Cette classification est en accord avec les données cliniques, tumorales et familiales collectées au sein du réseau français des laboratoires d'oncogénétique pour les patients porteurs de ces variations.

L'ensemble de ces travaux souligne l'importance des approches expérimentales combinées, à la fois au niveau de l'ARN mais également au niveau protéique, pour aboutir à une classification fiable des variations splicéogéniques à l'origine d'altérations en phase.

L'ensemble de ces travaux fait l'objet d'un article en cours de préparation pour publication.

Functional characterization of *MSH2* variants resulting into in-frame splicing alterations

Laëtitia Meulemans¹, Marie-Christine Waill², Stéphanie Baert-Desurmont^{1,3}, Gaia Castelain¹, Thierry Frebourg^{1,3}, Florence Coulet^{2,4}, Alexandra Martins¹, Martine Muleris^{2,4}, Pascaline Gaildrat^{1,*}

¹ Normandie Univ, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

² Department of Genetics, AP-HP, Hôpital Universitaire Pitié-Salpêtrière, Paris, France

³ Department of Genetics, Rouen University Hospital, Rouen, France

⁴ Sorbonne Université, Inserm, Centre de Recherche Saint-Antoine, CRSA, Paris, France

*To whom correspondence should be addressed. Email: pascaline.gaildrat@univ-rouen.fr

Conflict of interest

The authors declare that they have no conflict of interest.

Abstract

Background Spliceogenic variants in disease-causing genes are often classified as pathogenic since most of them induce frameshift RNA alterations resulting into protein loss-of-function. However, it is possible that certain variants can also cause in-frame splicing anomalies causing protein internal deletions or insertions that may preserve function, as recently shown for a subset of *BRCA2* variants. Here, we addressed this question by using as a model system *MSH2*, a DNA mismatch repair gene implicated in Lynch syndrome.

Methods First, eighteen *MSH2* variants, mostly localized within canonical splice sites, were fully characterized by using *in silico* prediction tools and splicing minigene assays. Next, the functionality of the resulting protein isoforms was assessed in a methylation tolerance-based assay.

Results Three specific biotypes were identified based on variant-induced in-frame spliceogenic outcomes: (i) exon skipping ($\Delta E3$, $\Delta E4$, $\Delta E5$, $\Delta E12$), (ii) segmental exonic deletions ($\Delta E7q48$, $\Delta E15p36$) and (iii) segmental intronic retentions ($\nabla E4p24$, $\nabla E7p9$, $\nabla E14p9$, $\nabla E12q30$). These splicing events were due to direct alterations of splice sites, as hinted by computational predictions. The resulting 10 *MSH2* protein isoforms exhibited either (i) large deletions (49 to 93 aa), (ii) small deletions (12 and 16 aa) or (iii) small insertions (3 to 10 aa) within different functional domains. All of these modifications abrogated *MSH2* function.

Conclusion Altogether, these data demonstrate that the analyzed *MSH2* in-frame spliceogenic variants do not lead to the production of functional protein isoforms, supporting their pathogenic classifications. This study stresses the importance of combining complementary approaches at both RNA and protein levels for accurate classification of in-frame spliceogenic variants.

Key words: In-frame splicing alterations, *MSH2* gene, Lynch syndrome, *in silico* predictions, minigene splicing assay, methylation tolerance assay, variant classification

Introduction

RNA splicing defects triggered by specific germline variants is a major cause of human hereditary monogenic disorders [1, 2]. Such spliceogenic impact is generally due to modifications of cis-acting splicing signals, notably consensus 3' and 5' splice sites (3'/5' ss) and/or of auxiliary splicing regulatory elements (SRE) [3–5]. The resulting consequence of these splicing code alterations is the production of aberrant transcripts with either lacking exonic portions or carrying intronic sequences. Most of these splicing anomalies lead to a frameshift of the reference open reading frame, which subsequently results into the introduction of a premature termination codon (PTC). This stop codon gain can then target the aberrant PTC-containing transcripts to selective degradation by a cellular surveillance pathway known as nonsense-mediated mRNA decay (NMD) [6, 7]. If the aberrant RNA escapes NMD, it can then be translated leading to the production of truncated proteins. Both situations generally cause the complete loss of functional protein. Based on this assumption, frameshift-inducing spliceogenic variants are systematically considered as null alleles and classified as pathogenic when they are detected in Mendelian disorder-causing genes in which loss-of-function (LoF) is an established disease mechanism [8].

However, certain spliceogenic variants can also trigger in-frame anomalies. In such circumstances, the possible outcomes at the protein level correspond to either an internal deletion or insertion of one to many amino acids (aa). The functional consequences of such modifications often remain difficult to apprehend and, consequently, the corresponding variants are usually classified as “variants of unknown significance” (VUS), hampering the clinical management of patients and their relatives [8]. It is also possible that some specific nucleotide changes initially classified as pathogenic based on their presumed LoF impacts would actually preserve protein function, at least partially, because of their in-frame splicing outcomes, as we recently demonstrated for certain variants in the breast and ovarian cancer predisposition *BRCA2* gene [9]. This could question the presumed LoF consequences of these variants and subsequently their initial pathogenic classifications [8, 9].

In this context, the combined assessments of variant-induced impact on both RNA splicing and protein function emerge as a prerequisite for accurately ascertaining the pathogenic or benign nature of such in-frame spliceogenic variants. In the present study, we addressed this question by using, as a model system, *MSH2* (MIM#609309), one of the four key DNA mismatch repair (MMR) genes implicated in the most prevalent hereditary cancer predispositions known as Lynch syndrome (LS, MIM#120435) [10–12]. Germline monoallelic LoF variants in this gene account for 33% of LS cases [13] and confer markedly increased lifetime risk of developing a spectrum of cancers, primarily colorectal and endometrial cancers, with age-related penetrance and variable expressivity [14–19]. Identification of pathogenic LoF mutations is critical for LS diagnosis and implementation of clinical strategies to reduce

risk of subsequent cancers in patients and for identification of at-risk pre-symptomatic relatives who could benefit from risk-reducing management [16, 20].

In this study, we first identified specific *MSH2* variants that lead to in-frame splicing anomalies. The nature (*i.e.* biotype) and the relative level of the variant-induced RNA alterations were assessed by using a cell-based minigene splicing assay and these experimental data were compared with *in silico* predictions. We then established the functionality of MSH2 protein isoforms presenting internal alterations corresponding to those induced by these spliceogenic variants by taking advantage of a recently developed MSH2 functional assay based on methylation tolerance assessment [21]. Altogether, these results should contribute to the comprehensive and accurate classification of in-frame *MSH2* spliceogenic variants.

Materials and Methods

Variant selection

Specific *MSH2* variants located at the invariant intronic dinucleotides within the 3' or 5' ss (AG/GT, hereafter termed IVS±1/2) of in-frame exons were retrieved from national and international human variation databases (Supplemental Materials and Methods). Additional *MSH2* variants were selected because they were identified as in-frame spliceogenic variants during the course of a systematic assessment of the impact on splicing of *MSH2* VUS identified in patients within the French consortium of oncogenetic laboratories for colorectal cancers (Unicancer, Cancer Genetic Group - GGC).

Bioinformatics predictions

Predictions of the variant impact on 3' or 5' ss were obtained by using the *in silico* algorithm MaxEntScan (MES) [22], interrogated via the Alamut Visual 2.11 integrated software (Interactive Biosoftware, SOPHiA GENETICS), setting the minimal MES score to zero.

Cell-based minigene splicing assay

In order to assess variant-induced impacts on splicing, we performed cell-based splicing assays by using the two-exon minigene pCAS2 vector, as previously described [9, 23, 24], with minor modifications (Supplementary Material and Methods, Supplementary Table S1). This assay is based on the comparative analysis of the splicing patterns of WT and mutant minigenes transiently expressed in HeLa cells.

Annotation of splicing events and associated terminology

The reference sequence used in the analysis of *MSH2* full-length transcript was: NM_000251.2. Normal exon inclusion is indicated by the symbol +E, followed by the number of the exon of interest. Based on previously published guidelines [25, 26], aberrant transcripts were annotated by using the following symbols: Δ , exonic deletion; \blacktriangledown , intronic retention; p, 3'ss shift; and q, 5'ss shift, followed by the number of nucleotides deleted or inserted. In this manuscript, the term “partial” is used to describe the deletion of a part of an exon or the retention of a part of an intron whereas the terms “incomplete” or “complete” refer to variant-induced splicing impacts with either the residual presence or the complete absence of the reference (full-length) transcript, respectively.

Methylation tolerance-based functional assay

In order to assess the functional consequences of the internal aa deletions or insertions resulting from *MSH2* variant-induced in-frame splicing anomalies, we took advantage on the recently developed methylation tolerance-based functional assay [21], with minor modifications (Supplementary Material and Methods).

Patient clinical, tumoral and family data

Clinical, tumoral and family data of patients carrying *MSH2* variants of interest were collected in collaboration with the French cancer genetics network (“Groupe Génétique et Cancer”, GGC, Unicancer). Individual written informed consent for genetic analysis was obtained from all participating patients.

Results

Characterization of *MSH2* variants responsible for in-frame splicing alterations

***MSH2* variants causing in-frame whole exon skipping**

We first selected all the exons which are in-frame in the *MSH2* gene. These correspond to exons 3, 4, 5 and 12 (Figure 1A). We then selected from human variation databases variants expected to induce in-frame skipping of these specific exons (Figure 1A, Table S2). This selection encompassed 7 presumed pathogenic variants, all located at the invariant IVS \pm 1/2 intronic dinucleotides and thus highly susceptible to disrupt either the physiological 3' or 5' ss, as indicated by the *in silico* predictions (Figures 1B, 1C, 1D, 1E).

In order to experimentally assess the impact on splicing of these specific variants, we took advantage of the cell-based pCAS2 minigene splicing assay. As expected, whereas the corresponding WT minigene constructs showed normal full exon inclusion (+E), all the 7 selected IVS \pm 1/2 variants were responsible

for drastic splicing alterations (Figures 1B, 1C, 1D and 1E). Indeed, for all variants tested, no +E transcripts were produced from the mutant minigenes. Two different splicing defects were observed depending on the variants. For 3 of them (c.645+1G>A, c.942+1_942+2del and c.1760-2A>G), their splicing impacts resulted into full exon skipping ($\Delta E3$, $\Delta E5$ or $\Delta E12 = 100\%$, respectively) (Figures 1B, 1D and 1E), whereas for the 4 remaining variants (c.792+1G>A, c.792+1G>C, c.2005+1G>A and c.2005+2T>C), the production of extra aberrant transcripts was detected in addition to exon skipping. More specifically, c.792+1G>A and c.792+1G>C caused a major exon skipping ($\Delta E4 = 93\%$ and 83% , respectively), as well as the minor production of an aberrant transcript with E4 inclusion followed by the retention of the first 38 nucleotides of intron 4 ($\nabla E4q38 = 7\%$ and 17% , respectively) (Figure 1C). For c.2005+1G>A and c.2005+2T>C, in addition to exon skipping ($\Delta E12 = 53\%$ and 43% , respectively), two abnormal transcripts were detected: one with the retention of the first 31 nucleotides of intron 12 ($\nabla E12q31 = 34\%$ and 51% , respectively) and the other with the deletion of the last 92 nucleotides of exon 12 ($\Delta E12q92 = 13\%$ and 6% , respectively) (Figure 1E). These extra aberrant transcripts result from the activation of cryptic exonic or intronic 5' ss, as indicated by *in silico* predictions (Figures 1C, 1E). Here, it is important to note that, in contrast to in-frame exon skipping ($\Delta E3$, $\Delta E4$, $\Delta E5$ and $\Delta E12$), all the splicing anomalies resulting from the use of cryptic 5' ss lead to frameshifts (Tables S3 and S4). This first part of the study allowed us to identify variants responsible for in-frame exon skipping, albeit to different extents. These splicing anomalies are susceptible to conduct to the production of MSH2 proteins with large internal deletions of 49 to 93 aa ($\Delta E3$, p.A123_Q215del; $\Delta E4$, p.I216_Q264del; $\Delta E5$, p.V265_Q314del; $\Delta E12$, p.Y588_G669del) (Tables S3 and S4).

MSH2 variants causing in-frame partial exonic deletion

Secondly, we extended our analysis to spliceogenic variants responsible for in-frame deletions of exonic portions (Figure 2A). These variants were identified during the course of a systematic assessment of the impact on splicing of MSH2 VUS reported in patients within the French network of oncogenetic laboratories (GGC).

In this context, we characterized the splicing impact of a missense variant, MSH2 c.1276G>A (p.G426R), located at the last base of exon 7. In the minigene-based splicing assay, this variant was responsible for a complete splicing anomaly resulting into the in-frame deletion of the last 48 nucleotides of exon 7 ($\Delta E7q48 = 100\%$) (Figure 2B). This splicing defect is the consequence of the variant-induced weakening of the physiological 5' ss strength which allows the use of an upstream cryptic exonic 5' ss, as indicated by MES predictions (Figure 2B). In parallel, we selected from mutational databases variants that were expected to exhibit the same splicing effect (Table S2, Sup Materials and Methods). In minigene-based assay, we demonstrated that c.1276+1G>A and c.1276+2T>A caused an in-frame

deletion ($\Delta E7q48 = 100\%$), identical to that observed with c.1276G>A, in good agreement with the MES predictions (Figure 2B).

The other variant identified as inducing an in-frame deletion of exonic portion is c.2459-3T>G. In the minigene-based assay, it caused an in-frame deletion of the first 36 nucleotides of exon 15 ($\Delta E15p36 = 52\%$), as well as two additional aberrant splicing events resulting into frame-shifts, *i.e.* exon skipping ($\Delta E15 = 43\%$) and a minor retention of the last 2 nucleotides of intron 14 ($\blacktriangledown E15p2 = 5\%$) (Figure 2C, Tables S3 and S4). These 3 aberrant transcripts are the consequence of the variant-induced weakening of the physiological 3'ss strength, the use of a downstream exonic cryptic 3'ss and the creation of an upstream intronic 3'ss, as indicated by the MES predictions (Figure 2C). In parallel, we selected from mutational databases variants that were expected to exhibit the same splicing impact (Table S2). We demonstrated that c.2459-2A>G and c.2459-1G>T caused different levels of in-frame partial exonic deletion ($\Delta E15p36 = 17\%$ and 71% , respectively) and of out-of-frame exon skipping ($\Delta E15 = 83\%$ and 29% , respectively), despite identical MES predictions (Figure 2C, Tables S3 and S4).

In this second part of the study, we characterized variants causing in-frame deletions of exonic portions, although to different degrees. These splicing alterations would potentially lead to the production of MSH2 protein isoforms with small internal deletions of 12 or 16 aa ($\Delta E15p36$, p.G820_A831del; $\Delta E7q48$, p.I411_G426del) (Tables S3 and S4).

MSH2 variants causing in-frame partial intronic retention

Finally, we expanded this study to spliceogenic variants causing in-frame intronic retentions (Figure 3A). The first 2 variants in this category are c.646-2A>G and c.646-1G>C which resulted in the retention of the last 27 and 24 nucleotides of intron 3 ($\blacktriangledown E4p27 = 55\%$ and 53% , $\blacktriangledown E4p24 = 45\%$ and 47% , respectively) (Figure 3B, Tables S2 and S3). These effects are the consequence of the variant-induced destruction of the physiological 3'ss concomitant to the use of two intronic cryptic 3'ss, as indicated by bioinformatics predictions (Figure 3B). Of note, the two splicing alterations are in-frame however the $\blacktriangledown E4p27$ event leads to the immediate introduction of a premature termination stop codon (PTC) (Table S4).

Another variant in this group is the VUS c.1077-11_1077-7del which caused in-frame retention of the last 9 nucleotides of intron 6 ($\blacktriangledown E7p9 = 90\%$), associated with a minor normal inclusion of exon 7 ($+E7 = 10\%$) (Figure 3C, Tables S2 and S3). This splicing impact resulted from the drastic variant-induced reduction in the strength of the physiological 3'ss and activation of an upstream cryptic intronic 3'ss, as predicted by *in silico* tools (Figure 3C).

The fourth variant identified in this class is c.2005+2del which led to in-frame retention of the first 30 nucleotides of intron 12 ($\blacktriangledown E12q30 = 71\%$) as well as in-frame skipping of exon 12 ($\Delta E12 = 25\%$) and a

very minor out-of-frame deletion of the last 92 nucleotides of exon 12 ($\Delta E12q92 = 4\%$) (Figure 3D, Tables S2 and S3). These splicing alterations are the result of the variant-induced abolition of the physiological 5' ss and the activation of a downstream cryptic intronic 5' ss, predicted by the MES algorithm (Figure 3D), and the use of an upstream exonic cryptic 5'ss, not predicted by the MES algorithm but predicted by the SSF-L algorithm (data not shown).

The last tested variant is c.2211-10T>G which mainly induced an in-frame insertion of the 9 last nucleotides of intron 13 ($\nabla E14p9 = 67\%$) and 3 additional minor out-of-frame splicing alterations corresponding to the skipping of exon 14 ($\Delta E14 = 18\%$) or the deletion of the first 169 or 200 nucleotides of this exon ($\Delta E14p169 = 6\%$, $\Delta E14p200 = 9\%$) (Figure 3E, Tables S2 and S3). The predominant production of the $\nabla E14p9$ isoform is due to the variant-induced destruction of the natural 3'ss concomitant to the creation of an upstream intronic 3'ss, whereas the $\Delta E4p169$ and $\Delta E4p200$ transcripts result from the activation of two downstream cryptic exonic 3'ss, as suggested by the *in silico* predictions (Figure 3E).

In this third part of the study, we described variants causing in-frame intronic retentions to variable levels. If translated, the corresponding RNA isoforms would lead to the production of MSH2 protein isoforms with small internal insertions of 3 to 10 aa ($\nabla E7p9$, p.R358_S359ins3; $\nabla E14p9$, p.R737_S738ins3; $\nabla E4p24$, p.E215_I216ins8; $\nabla E12q30$, p.G669_P670ins10) (Tables S3 and S4).

Functional analysis of MSH2 protein isoforms resulting from variant-induced in-frame splicing alterations

On the basis of the variant-induced in-frame spliceogenic outcomes, 10 MSH2 protein isoforms were defined with either (i) large internal deletions (49 to 93 aa) ($\Delta E3$, p.A123_Q215del; $\Delta E4$, p.I216_Q264del; $\Delta E5$, p.V265_Q314del; $\Delta E12$, p.Y588_G669del), (ii) small internal deletion (12 and 16 aa) ($\Delta E15p36$, p.G820_A831del; $\Delta E7q48$, p.I411_G426del) or (iii) small internal insertions (3 to 10 aa) ($\nabla E7p9$, p.R358_S359ins3; $\nabla E14p9$, p.R737_S738ins3; $\nabla E4p24$, p.E215_I216ins8; $\nabla E12q30$, p.G669_P670ins10) within different functional domains (Figure 4 and Table S4). In order to assess the functional consequences of the spliceogenic in-frame variants responsible for the production of these different protein isoforms, we used the recently published methylation tolerance-based functional assay [21]. Briefly, this test relies on the ability of the MMR system to trigger apoptotic signaling in the presence of DNA damage induced by methylating agents. Deficient MMR cells do not recognize such lesions and are therefore able to escape apoptosis (they exhibit the so-called methylation tolerance phenotype). The assay evaluates the ability of the in-frame isoforms expressed in LoVo cells, a *MSH2*-deficient cell line, to complement the methylation tolerance phenotype. We demonstrated that all the in-frame deletions or insertions tested were unable to restore the MMR-dependent DNA damage

signalling (Figure 5). Still, one should note that one variant (∇ E14p9, p.R737_S738ins3) exhibited a slightly lower mean survival score, as compared to known pathogenic missense variants. In contrast, control neutral missense variants were active.

Patient clinical, tumoral and family data

Clinical phenotype, tumoral data and family history were collected within the French GGC network for 40 patients carrying 16 different *MSH2* variants among the 18 in-frame spliceogenic variants described in this study. All the information are summarized in Supplemental Table S5. The most frequent cancers reported in patients were colorectal (CRC) and endometrium (EC) cancers with a median age at onset of 49 year-old (range 26 - 77) and 50 year-old (range 40 - 75), respectively. All tested tumors from variant carriers but 1 were MSI and/or exhibited loss of MSH2 protein expression. Overall, these patient clinical, tumoral and family data are supportive of the pathogenic classification of the *MSH2* variants selected in this study.

Discussion

In this study, we have provided the first functional characterization of in-frame spliceogenic *IVS \pm 1/2* variants in a key MMR gene. By demonstrating that none of the resulting protein indel isoforms retain function, we ascertained the pathogenic nature of these specific variants. Generally, RNA analyzes are employed to demonstrate variant-induced splicing modifications resulting into frameshift [27–29], whereas functional protein assessments are mainly conducted for evaluating the effect of missense variants [21, 30–32]. So far, little attention has been paid to MMR variants potentially responsible for in-frame modifications through splicing defects, while their functional and clinical consequences may potentially range from disease-causing LoF to neutral impact. Here, we especially focused on *MSH2* *IVS \pm 1/2* variants susceptible to have such effects. Within the current consensual framework for variant classification [8, 33], these variants are *a priori* linked to the strongest type of evidence in favor of pathogenicity (PVS1), as they are assumed to trigger LoF. However, caution in the systematic application of this criteria are recommended since some *IVS \pm 1/2* variants can lead to in-frame splicing alterations that potentially allow the preservation of protein function, at least partially. This has been well illustrated by recent works on specific variants in one of the major predisposing breast and ovarian cancer gene, *BRCA2*. Indeed, we and others demonstrated that certain *BRCA2* *IVS \pm 1/2* variants, as well as nonsense mutations, trigger in-frame splicing modifications conducting to the production of protein isoforms deleted for specific exons but still partially functional [9, 34, 35]. The potential existence of such rescue mechanism highlights the need to characterize the biological impact of potential in-frame

spliceogenic variants at both RNA and protein levels in order to ensure their accurate and comprehensive classification.

Here, we addressed this question by using as a model system another major hereditary cancer gene, *MSH2*. In this study, by assessing the splicing impact of 18 variants including 14 IVS±1/2, we characterized 3 distinct biotypes of in-frame spliceogenic *MSH2* variants, on the basis of their RNA and protein outcomes, corresponding to either (i) full exon skipping resulting into large protein internal deletions (del 49 to 93 aa), (ii) exonic segmental deletion resulting into protein small internal deletions (del 12 and 16 aa) or (iii) intronic segmental retention resulting into protein small internal insertions (ins 3 to 10 aa) (Figure S1).

All these in-frame mis-splicing events are the consequences of variant-induced disruptions of the physiological 5'/3'ss, alone or combination with to the activation of cryptic exonic or intronic 5'/3'ss. These observations were in good agreement with *in silico* MES predictions, confirming that this bioinformatics approach is a reliable prediction tool to anticipate such impacts, as previously shown [36, 37]. However, such *in silico* approach does not allow foreseeing the specific levels of potential cryptic 5'/3'ss usage. Indeed, we showed here that two IVS±1/2 variants (*i.e.* *MSH2* c.2459-1G>T and c.2459-2A>G) induced either major or minor usage of an in-frame cryptic 3'ss, while associated with identical MES predictions. These results suggest a dual distinct variant effect on both splice sites and SRE, as already hypothesized for certain *BRCA2* variants [38, 39]. Based on these observations and in absence of reliable prediction tool to anticipate such concomitant effects, we recommend to experimentally test for their splicing impact all variants, even IVS±1/2, especially if a cryptic 5'/3'ss is predicted in the vicinity of the nucleotide changes.

In this study, the minigene-based assay, in agreement with patient RNA data, allowed not only to accurately characterize the nature of the different RNA in-frame and frameshift isoforms produced specifically in the mutant context, but it also permitted to precisely quantify their relative proportions. Among the 18 selected variants, 12 of them induced the production of multiple aberrant in-frame or frameshift transcripts, with up to 4 distinct RNA isoforms identified for one specific variant. These observations underline the importance of the exhaustive and quantitative assessment of mis-splicing for accurately infer the variant biological impact. Such information are not always available from patient RNA analyses. Still, it is to note that some of the variant-induced in-frame mis-splicing detected by using minigene assays in this study have been previously detected in patient RNA samples (Table S6), confirming the biological relevance of the minigene splicing reporter system, as previously shown for MMR genes [28, 40].

The strength of our strategy was to combine RNA analyses and protein functional assays to ascertain the comprehensive biological impact of potential in-frame spliceogenic variants. The methylation tolerance-based functional assay used in this study examined one of the two key MMR functions by measuring the cellular ability to instigate apoptosis in response to DNA damage [41, 42]. Recent work demonstrated that this approach exhibits 95% sensitivity and 100% specificity in classifying *MSH2* and *MLH1* missense variants as pathogenic or neutral [21], indicating that this well-calibrated system is a pertinent method to gain insight into variant interpretation. We took advantage of this approach to assess the functional consequences of the different in-frame spliceogenic variants and showed that all the corresponding protein indel isoforms resulted in LoF. Still, it is to note that one of these protein isoforms (*i.e.* p.R737_S738ins3 resulting from the spliceogenic *MSH2* 2211-10T>G variant) displays a mean survival score that is slightly lower as compared to the known pathogenic variants of the calibration set. Further analyses will be required to determine if this specific variant is responsible for total loss of function.

Altogether, these data demonstrated that *MSH2* function does not tolerate any of the deletions/insertions triggered by these spliceogenic variants. They confirmed the functional relevance of the *MSH2* modified sequences, in good agreement with previous structural and functional studies mapping essential domains within this protein [43–46]. To our knowledge, only two of these 10 protein isoforms (*i.e.* p.V265_Q314del, ΔE5 and p.E215_I216ins8, ▼E4p24) have been previously tested in a functional assay based on stable murine/human hybrid cell lines, with concordant results (*i.e.* defective MMR activity) [47]. Indirectly, it was already known that ΔE5 protein isoform was dysfunctional as the most frequent of all known pathogenic *MSH2* variants (*i.e.* c.942+3A>T) is indeed a spliceogenic variant leading to the in-frame skipping of this specific exon [48]. In addition, it should be pointed out that the ΔE5 as well as ΔE3 RNA have been previously reported as naturally occurring alternative transcripts [26]. Because we demonstrated that these specific in-frame splicing events do not lead to *MSH2* functional protein, we can now exclude the possibility that not only the ΔE5 but also the ΔE3 alternative transcripts represent “rescue” isoforms, in contrast to what has been suggested for *BRCA1* ΔE9-10 [49].

The loss of function caused by the in-frame spliceogenic variants could be the consequence of the alteration of specific *MSH2* essential functional domains [43–46], as each of them might need to be fully intact to be active. Under this hypothesis, defective MMR caused by *MSH2* spliceogenic in-frame mutations would be the result of the failure of one of the multiple specific activities supporting *MSH2* function, such as protein-protein interaction (*MSH2*–*MSH6* complex formation), DNA binding (mismatch recognition) or enzymatic activity (ATP binding/hydrolysis). Another explanation would be that these in-frame modifications impaired *MSH2* function by conferring a defect in protein folding

and structural integrity leading eventually to lower activity or to protein destabilization and degradation. This second hypothesis might be favored because it has been previously shown that 2 of the 10 in-frame isoforms tested here (*i.e.* ΔE5, p.V265_Q314del and ▼E4p24, p.E215_I216ins8) lead to the absence of MSH2 protein detection in murine/human hybrid cell lines [47]. In addition, it has been previously reported that many missense pathogenic *MSH2* variants also impair MMR function through a defect in protein stability [41, 50].

Altogether, our data demonstrate that MSH2 function does not tolerate any of the in-frame alterations caused by the spliceogenic variants characterized in this study. Still, we cannot completely exclude the possibility that other *MSH2* SNVs, or genomic deletions or insertions allow the production of additional different internally or truncated protein isoforms that may conserve activity. Recently, it has been shown by using deep mutational scanning combined with massively parallel functional MMR assay that MSH2 exhibits a remarkable degree of tolerance to missense variants [32]. Indeed, despite the high degree of evolutionary sequence conservation, the large majority (89%) of the ~ 17 000 assayed single amino acid substitutions within this protein do not lead to LoF. By providing the first exhaustive cDNA-based functional effect map of MSH2 missense variants, this study revealed in particular that the MSH2 N-terminal sequence, coding for the region referred to as mismatch binding domain (residues 1-124), is largely devoided of missense LoF variants. These results are in agreement with the demonstrated dispensability of this region in the yeast MSH2 orthologue [51] and with the observation that the variant-induced translation start codon loss (*MSH2* c.1A>C) lead to the production of a protein deleted for its first 25 aa but still partly functional, which may be responsible for the reduced penetrance of this variant [52–55]. Based on these data, it is conceivable that certain deletions within the N-terminal region of *MSH2* may preserve protein functionality.

Altogether, our results contribute to the clinical classification of the *MSH2* in-frame spliceogenic variants characterized in this study. By ascertaining their LoF status, we demonstrated that none of the *MSH2* in-frame spliceogenic IVS±1/2 variants retain/sustain activity. This contrasts with *BRCA2*, another hereditary cancer gene, where certain of this type of variants don't lead to the total abrogation of tumor-suppressor activity, resulting into the questioning of their pathogenic classification [9, 34, 35]. Conversely, our data support the initial pathogenic classification of the 14 in-frame spliceogenic IVS±1/2 variants, consistent with the collected/respective patient clinical, tumoral and family data. Moreover, this interpretation can be extended to all *MSH2* genomic deletions reported in mutational databases as responsible for the same types of in-frame modifications. Our data also support the classification of 4 additional in-frame spliceogenic variants located outside IVS±1/2 (3 intronic and 1 missense) as pathogenic.

The strength strategy was to combine RNA analyses and protein functional assays to accurately ascertain the overall biological impact of potential in-frame spliceogenic variants. Our study framework required to sequentially evaluate the variant-induced splicing modifications and their protein functional consequences into two separate independent assays. The functional assay used here relies on the cDNA-based expression of the mutant protein isoforms which is not appropriate for the assessment of the variant effect on splicing. It would be interesting in the future to gain insight into the variant impact on both RNA splicing and protein function in a unique integrated system. Notably, one could take advantage of a functional assay which relies on the introduction of the variants into the endogenous *MSH2* chromosomal locus by using method such as CRISPR-Cas9 gene editing, as recently described for *MSH2* missense variants [31].

Acknowledgements

The authors thank all the clinical and molecular geneticists of the French oncogenetic network (GGC). We are also grateful to the Flow Cytometry core (CYTO-ICAN) of ICAN institute for the flow cytometry analysis. This work was financially supported by the Fédération Hospitalo-Universitaire (FHU) Normandy Centre for Genomic and Personalized Medicine (NGP), the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (Gefluc, # R18064EE), a translational research grant from the French National Cancer Institute and the Direction Generale de l'Offre des Soins (INCa/DGOS, AAP/CFB/CI), as well as the European Union and Région Normandie. Europe gets involved in Normandie with the European Regional Development Fund (ERDF). L. Meulemans was funded by a FHU-NGP PhD fellowship and benefit of two short-term fellowship from Cancéropôle Nord-Ouest (CNO) and EDnBISE for mobility.

References

- 1 Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016;17:19–32.
- 2 Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, Bayrak-Toydemir P, McDonald J, Fairbrother WG. Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* 2017;49:848–55.
- 3 Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 2002;3:285–98.
- 4 Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat* 2007;28:150–8.
- 5 Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res* 2011;21:1563–71.
- 6 Popp MW-L, Maquat LE. Organizing Principles of Mammalian Nonsense-Mediated mRNA Decay. *Annu Rev Genet* 2013;47:139–65.
- 7 Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci* 2016;129:461–7.
- 8 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehml HL, on behalf of the ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–23.
- 9 Meulemans L, Mesman RLS, Caputo SM, Krieger S, Guillaud-Bataille M, Caux-Moncoutier V, Léone M, Boutry-Kryza N, Sokolowska J, Révillion F, Delnatte C, Tubeuf H, Soukarieh O, Bonnet-Dorion F, Guibert V, Bronner M, Bourdon V, Lizard S, Vilquin P, Privat M, Drouet A, Grout C, Calléja FMGR, Golmard L, Vrieling H, Stoppa-Lyonnet D, Houdayer C, Frebourg T, Vreeswijk MPG, Martins A, Gaildrat P. Skipping Nonsense to Maintain Function: The Paradigm of *BRCA2* Exon 12. *Cancer Res* 2020;80:1374–86.
- 10 Lynch HT, de la Chapelle A. Hereditary Colorectal Cancer. *N Engl J Med* 2003;48:919–932.
- 11 Lynch H, Lynch P, Lanspa S, Snyder C, Lynch J, Boland C. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 2009;76:1–18.
- 12 Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895–2015. *Nat Rev Cancer* 2015;15:181–94.
- 13 Plazzer JP, Sijmons RH, Woods MO, Peltomäki P, Thompson B, Den Dunnen JT, Macrae F. The InSiGHT database: utilizing 100 years of insights into Lynch Syndrome. *Fam Cancer* 2013;12:175–80.
- 14 Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, Ki PP, Mecklin J-P, Rvinen HJJ. Cancer risk in mutation carriers of DNA-mismatch-repair genes. *Int. J. Cancer* 1999;81:214–8.
- 15 Hampel H, Khanduja K, Prior TW, Comeras I. Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). *N Engl J Med* 2005;352:1851–1860.

- 16 Järvinen HJ, Renkonen-Sinisalo L, Aktán-Collán K, Peltomäki P, Aaltonen LA, Mecklin J-P. Ten Years After Mutation Testing for Lynch Syndrome: Cancer Incidence and Outcome in Mutation-Positive and Mutation-Negative Family Members. *J Clin Oncol* 2009;27:4793–7.
- 17 Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, Lindblom A, Macrae F, Blanco I, Sijmons RH, Jeffries J, Vasen HFA, Burn J, Nakken S, Hovig E, Rødland EA, Tharmaratnam K, de Vos tot Nederveen Cappel WH, Hill J, Wijnen JT, Jenkins MA, Green K, Laloo F, Sunde L, Mints M, Bertario L, Pineda M, Navarro M, Morak M, Renkonen-Sinisalo L, Valentin MD, Frayling IM, Plazzer J-P, Pylvanainen K, Genuardi M, Mecklin J-P, Moeslein G, Sampson JR, Capella G. Cancer risk and survival in *path_MMR* carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut* 2018;67:1306–16.
- 18 ten Broeke SW, van Bavel TC, Jansen AML, Gómez-García E, Hes FJ, van Hest LP, Letteboer TGW, Olderode-Berends MJW, Ruano D, Spruijt L, Suerink M, Tops CM, van Eijk R, Morreau H, van Wezel T, Nielsen M. Molecular Background of Colorectal Tumors From Patients With Lynch Syndrome Associated With Germline Variants in PMS2. *Gastroenterology* 2018;155:844–51.
- 19 Dominguez-Valentin M, Sampson JR, Seppälä TT, ten Broeke SW, Plazzer J-P, Nakken S, Engel C, Aretz S, Jenkins MA, Sunde L, Bernstein I, Capella G, Balaguer F, Thomas H, Evans DG, Burn J, Greenblatt M, Hovig E, de Vos tot Nederveen Cappel WH, Sijmons RH, Bertario L, Tibiletti MG, Cavestro GM, Lindblom A, Della Valle A, Lopez-Köstner F, Gluck N, Katz LH, Heinemann K, Vaccaro CA, Büttner R, Görgens H, Holinski-Feder E, Morak M, Holzapfel S, Hüneburg R, Knebel Doeberitz M von, Loeffler M, Rahner N, Schackert HK, Steinke-Lange V, Schmiegel W, Vangala D, Pylvänäinen K, Renkonen-Sinisalo L, Hopper JL, Win AK, Haile RW, Lindor NM, Gallinger S, Le Marchand L, Newcomb PA, Figueiredo JC, Thibodeau SN, Wadt K, Therkildsen C, Okkels H, Ketabi Z, Moreira L, Sánchez A, Serra-Burriel M, Pineda M, Navarro M, Blanco I, Green K, Laloo F, Crosbie EJ, Hill J, Denton OG, Frayling IM, Rødland EA, Vasen H, Mints M, Neffa F, Esperon P, Alvarez K, Kariv R, Rosner G, Pinero TA, Gonzalez ML, Kalfayan P, Tjandra D, Winship IM, Macrae F, Mösllein G, Mecklin J-P, Nielsen M, Møller P. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. *Genet Med* 2020;22:15–25.
- 20 Vasen HFA, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, Bernstein I, Bertario L, Burn J, Capella G, Colas C, Engel C, Frayling IM, Genuardi M, Heinemann K, Hes FJ, Hodgson SV, Karagiannis JA, Laloo F, Lindblom A, Mecklin J-P, Møller P, Myrhoj T, Nagengast FM, Parc Y, Ponz de Leon M, Renkonen-Sinisalo L, Sampson JR, Stormorken A, Sijmons RH, Tejpar S, Thomas HJW, Rahner N, Wijnen JT, Järvinen HJ, Mösllein G. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut* 2013;62:812–23.
- 21 Bouvet D, Bodo S, Munier A, Guillerme E, Bertrand R, Colas C, Duval A, Coulet F, Muleris M. Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the MLH1 and MSH2 Genes and Identify Patients With Lynch Syndrome. *Gastroenterology* 2019;157:421–31.
- 22 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.
- 23 Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* 2010;653:249–57.

- 24 Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet* 2016;12:e1005756.
- 25 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, Brown M, Blanco A, Bonache S, Menéndez M, Houdayer C, Foglia C, Fackenthal JD, Baralle D, Wappenschmidt B, kConFaB Investigators, Díaz-Rubio E, Caldés T, Walker L, Díez O, Vega A, Spurdle AB, Radice P, De La Hoya M. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum Mol Genet* 2014;23:3666–80.
- 26 Thompson BA, Martins A, Spurdle AB. A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin Genet* 2015;87:100–8.
- 27 Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, Puisieux A, Wang Q. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat* 2006;27:145–54.
- 28 Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine MP, Soret J, Tazi J, Frébourg T, Tosi M. A large fraction of unclassified variants of the mismatch repair genes *MLH1* and *MSH2* is associated with splicing defects. *Hum Mutat* 2008;29:1412–24.
- 29 Thompson BA, Walters R, Parsons MT, Dumenil T, Drost M, Tiersma Y, Lindor NM, Tavtigian SV, de Wind N, Spurdle AB, the InSiGHT Variant Interpretation Committee. Contribution of mRNA Splicing to Mismatch Repair Gene Sequence Variant Interpretation. *Front Genet* 2020;11:798.
- 30 Drost M, Tiersma Y, Thompson BA, Frederiksen JH, Keijzers G, Glubb D, Kathe S, Osinga J, Westers H, Pappas L, Boucher KM, Molenkamp S, Zonneveld JB, van Asperen CJ, Goldgar DE, Wallace SS, Sijmons RH, Spurdle AB, Rasmussen LJ, Greenblatt MS, de Wind N, Tavtigian SV. A functional assay-based procedure to classify mismatch repair gene variants in Lynch syndrome. *Genet Med* 2019;21:1486–96.
- 31 Rath A, Mishra A, Ferreira VD, Hu C, Omerza G, Kelly K, Hesse A, Reddi HV, Grady JP, Heinen CD. Functional interrogation of Lynch syndrome-associated *MSH2* missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Hum Mutat* 2019;40:2044–56.
- 32 Jia X, Burugula BB, Chen V, Lemons RM, Jayakody S, Maksutova M, Kitzman JO. Massively parallel functional testing of *MSH2* missense variants conferring Lynch syndrome risk. *Am J Hum Genet* 2021;108:163–75.
- 33 Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018;39:1517–24.
- 34 Stauffer S, Biswas K, Sharan SK. Bypass of premature stop codons and generation of functional BRCA2 by exon skipping. *J Hum Genet* 2020;65:805–9.
- 35 Mesman RLS, Calléja FMGR, de la Hoya M, Devilee P, van Asperen CJ, Vrieling H, Vreeswijk MPG. Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. *Genet Med* 2020;22:1355–65.
- 36 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, Mazoyer S, Muller D, Remenieras A, Révillion F,

- Rouleau E, Sokolowska J, Vert J-P, Lidereau R, Soubrier F, Sobol H, Sevenet N, Bressac-de Paillerets B, Hardouin A, Tosi M, Sinilnikova OM, Stoppa-Lyonnet D. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 2012;33:1228–38.
- 37 Shamsani J, Kazakoff SH, Armean IM, McLaren W, Parsons MT, Thompson BA, O'Mara TA, Hunt SE, Waddell N, Spurdle AB. A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinformatics* 2019;35:2315–7.
- 38 Canson D, Glubb D, Spurdle AB. Variant effect on splicing regulatory elements, branchpoint usage, and pseudoexonization: Strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Hum Mutat* 2020;41:1705–21.
- 39 Kováčová T, Souček P, Hujová P, Freiburger T, Grodecká L. Splicing Enhancers at Intron–Exon Borders Participate in Acceptor Splice Sites Recognition. *Int J Mol Sci* 2020;21:6553.
- 40 van der Klift HM, Jansen AML, van der Steenstraten N, Bik EC, Tops CMJ, Devilee P, Wijnen JT. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med* 2015;3:327–45.
- 41 Heinen CD. Mismatch repair defects and Lynch syndrome: The role of the basic scientist in the battle against cancer. *DNA Repair* 2016;38:127–34.
- 42 Li Z, Pearlman AH, Hsieh P. DNA mismatch repair and the DNA damage response. *DNA Repair* 2016;38:94–101.
- 43 Lamers MH, Perrakis A, Enzlin JH, Winterwerp HHK, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G-T mismatch. *Nature* 2000;407:711–7.
- 44 Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. Structure of the Human MutS α DNA Lesion Recognition Complex. *Mol Cell* 2007;26:579–92.
- 45 Lützen A, de Wind N, Georgijevic D, Nielsen FC, Rasmussen LJ. Functional analysis of HNPCC-related missense mutations in MSH2. *Mutat Res* 2008;645:44–55.
- 46 Edelbrock MA, Kaliyaperumal S, Williams KJ. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutat Res* 2013;743–744:53–66.
- 47 Marra G, D'Atri S, Yan H, Perrera C, Cannavo' E, Vogelstein B, Jiricny J. Phenotypic Analysis of hMSH2 Mutations in Mouse Cells Carrying Human Chromosomes. *Cancer Res* 2001;61:7719–21.
- 48 Desai DC, Lockman JC, Chadwick RB, Gao X, Percesepe A, Evans DG, Miyaki M, Yuen ST, Radice P, Maher ER, Wright FA, de La Chapelle A. Recurrent germline mutation in MSH2 arises frequently de novo. *J Med Genet* 2000;37:646–52.
- 49 de la Hoya M, Soukarieh O, López-Perolio I, Vega A, Walker LC, van Ierland Y, Baralle D, Santamariña M, Lattimore V, Wijnen J, Whiley P, Blanco A, Raponi M, Hauke J, Wappenschmidt B, Becker A, Hansen TVO, Behar R, Investigators Kc, Niederacher D, Arnold N, Dworniczak B, Steinemann D, Faust U, Rubinstein W, Hulick PJ, Houdayer C, Caputo SM, Castera L, Pesaran T, Chao E, Brewer C, Southey MC, van Asperen CJ, Singer CF, Sullivan J, Poplawski N, Mai P, Peto J, Johnson N, Burwinkel B, Surowy H, Bojesen SE, Flyger H, Lindblom A, Margolin S, Chang-Claude J, Rudolph A, Radice P, Galastri L, Olson JE, Hallberg E, Giles GG, Milne RL, Andrulis IL, Glendon G, Hall P, Czene K, Blows F, Shah M,

Wang Q, Dennis J, Michailidou K, McGuffog L, Bolla MK, Antoniou AC, Easton DF, Couch FJ, Tavtigian S, Vreeswijk MP, Parsons M, Meeks HD, Martins A, Goldgar DE, Spurdle AB. Combined genetic and splicing analysis of BRCA1 c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet* 2016;25:2256–68.

50 Heinen CD, Rasmussen L. Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. *Hered Cancer Clin Pract* 2012;10:9.

51 Kumar C, Piacente SC, Sibert J, Bukata AR, O'Connor J, Alani E, Surtees JA. Multiple Factors Insulate Msh2–Msh6 Mismatch Repair Activity from Defects in Msh2 Domain I. *J Mol Biol* 2011;411:765–80.

52 Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, Williams N, Warner J, Campbell H, Porteous ME, Dunlop MG. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. *Hum Mutat* 2008;29:367–74.

53 Kets CM, Hoogerbrugge N, van Krieken JHJM, Goossens M, Brunner HG, Ligtenberg MJL. Compound heterozygosity for two MSH2 mutations suggests mild consequences of the initiation codon variant c.1A>G of MSH2. *Eur J Hum Genet* 2009;17:159–64.

54 Cyr JL, Brown GD, Stroop J, Heinen CD. The predicted truncation from a cancer-associated variant of the *MSH2* initiation codon alters activity of the MSH2-MSH6 mismatch repair complex. *Mol Carcinog* 2012;51:647–58.

55 Rosenthal ET, Bowles KR, Pruss D, van Kan A, Vail PJ, McElroy H, Wenstrup RJ. Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin Genet* 2015;88:533–41.

Legends to Figures

Figure 1. Spliceogenic *MSH2* variants causing in-frame whole exon skipping

A. Distribution of the 7 selected variants. The figure shows the schematic representation of the *MSH2* gene with the phasing of its 16 exons. The exons which are in-frame are indicated in black. The variants of interest are located at the canonical intronic dinucleotides (IVS±1/2) within the 3' or 5' ss of these specific exons. **B-E. Variant-induced impacts on splicing of exon 3 (B), exon 4 (C), exon 5 (D) and exon 12 (E), as determined in a cell-based minigene splicing assay.** Wild-type (WT) and mutant pCAS2 minigene constructs for the exon of interest were transiently expressed in HeLa cells and the splicing patterns of the transcripts produced from the different minigenes were analyzed by semi-quantitative fluorescent RT-PCR, as indicated in Materials and Methods. The relative levels of the different RNA isoforms were quantified by resolving the fluorescent RT-PCR products by capillary electrophoresis under denaturing conditions on an automated 3500 Genetic Analyzer (Applied Biosystems). One representative electropherogram analyzed by using the GeneMapper v5.0 Software (Applied Biosystems) is shown for each variant and WT constructs. The fluorescence peaks correspond to the different RT-PCR products resolved according to their respective sizes. The identities of the different transcripts are indicated on the top of each panel: transcripts with normal exon inclusion (+E), transcripts with whole or partial deletions (Δ E) and transcripts with intronic retentions (∇ E). The symbols p and q represent respectively 3' and 5' ss shift, followed by the number of nucleotides deleted or inserted. The RNA isoforms indicated in black are in-frame whereas those reported in grey are out-of-frame. Peak areas were used to quantify the relative levels of each transcript, expressed as percentages, and correspond to the average of at least three independent experiments. On the left, the splicing patterns of the physiological and aberrant transcripts are schematically represented. On the right, the *in silico* MES predictions of the physiological (Ψ), new (n) or cryptic (c) 3'/5' ss are shown in the WT and variant contexts (Alamut Visual 2.11, Interactive Biosoftware).

Figure 2. Spliceogenic *MSH2* variants causing in-frame partial exonic deletion

A. Distribution of the 6 selected variants. The figure shows the schematic representation of the *MSH2* gene with the phasing of its 16 exons. The variants of interest are located at the 3' or 5' ss of out-of-frame exons indicated in dark grey. **B-C. Variant-induced impacts on splicing of exon 7 (B) and exon 15 (C), as determined in cell-based minigene splicing assays.** Legends as described in Figure 1. B-E. When transcripts were close in size, stars (*) were used to link RNA isoforms with their corresponding relative levels.

Figure 3. Spliceogenic *MSH2* variants causing in-frame partial intronic retention

A. Distribution of the 5 selected variants. The figure shows the schematic representation of the *MSH2* gene with the phasing of its 16 exons. The variants of interest are located at the 3' or 5' ss of in-frame or out-of-frame exons indicated, respectively, in black and dark grey. **B-E. Variant-induced impacts on splicing of exon 4 (B), exon 7 (C), exon 12 (D) and exon 14 (E), as determined in cell-based minigene splicing assays.** Legends as described in Figure 1. B-E.

Figure 4. Schematic representation of the variant-induced in-frame splicing modifications at the RNA level and the corresponding consequences at the MSH2 protein level.

The top panel shows the schematic representation of the *MSH2* mRNA with its 16 exons. In-frame exons are indicated in black. The different variant-induced in-frame splicing anomalies are indicated, corresponding to either (i) exon skipping (ΔE), (ii) partial exonic deletion ($\Delta E_{p/q}$) or (iii) partial intronic retention ($\nabla E_{p/q}$). The symbols p and q represent 3' and 5' ss shift, respectively, followed by the number of nucleotides deleted or inserted. The bottom panel shows the schematic representation of the MSH2 protein with its different functional domains, as reported by Lutzen *et al*, 2008. The respective consequences at the protein level of the different in-frame splicing anomalies are indicated, corresponding to either (i) large internal deletions (49 to 93 aa, 4 protein isoforms), (ii) small internal deletion (12 and 16 aa, 2 protein isoforms) or (iii) small internal insertions (3 to 10 aa, 4 protein isoforms) within different functional domains.

Figure 5. Functional analysis of MSH2 protein isoforms resulting from variant-induced in-frame splicing alterations

A total of 10 MSH2 protein isoforms were tested in methylation tolerance-based functional assays, in parallel to the WT protein and to known pathogenic (n=10) or neutral (n=10) missense variants, used as controls. The 10 MSH2 protein isoforms resulting from variant-induced in-frame splicing alterations can be divided into 3 distinct biotypes: (i) exon skipping resulting into protein large internal deletions (n=4), (ii) partial exonic deletions resulting into protein small internal deletions (n=2) or (iii) partial intronic retentions resulting into protein small internal insertions (n=4). The mean survival score represents the average survival fraction after 1 and 2 treatments with 1 μ M MNNG. The *gray area* corresponds to the 5% values that differed significantly from the average as defined by the z-score. The cutoff value as determined by the ROC curve in the validation set is represented by the *dotted line*.

Figure 1

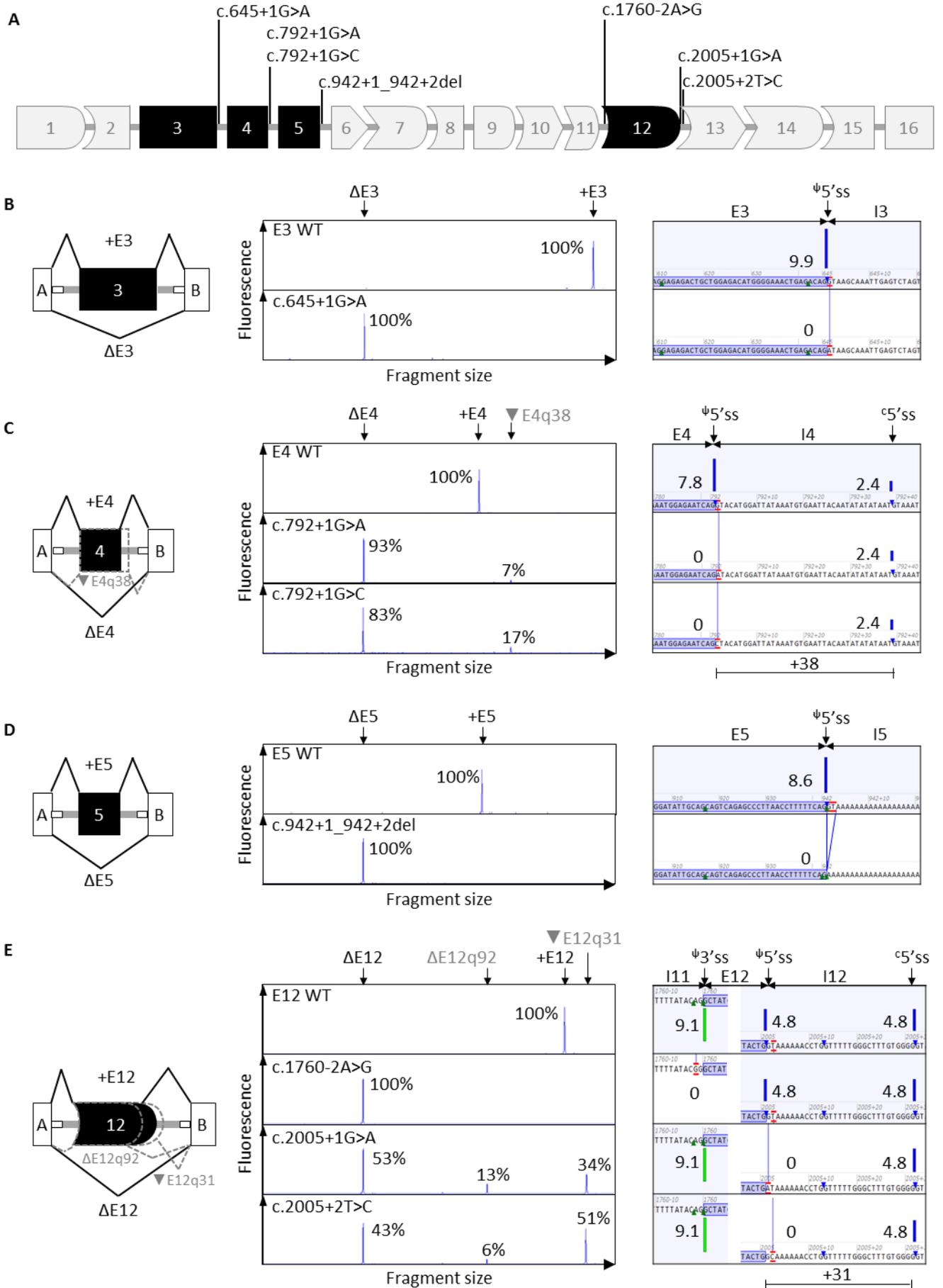


Figure 2

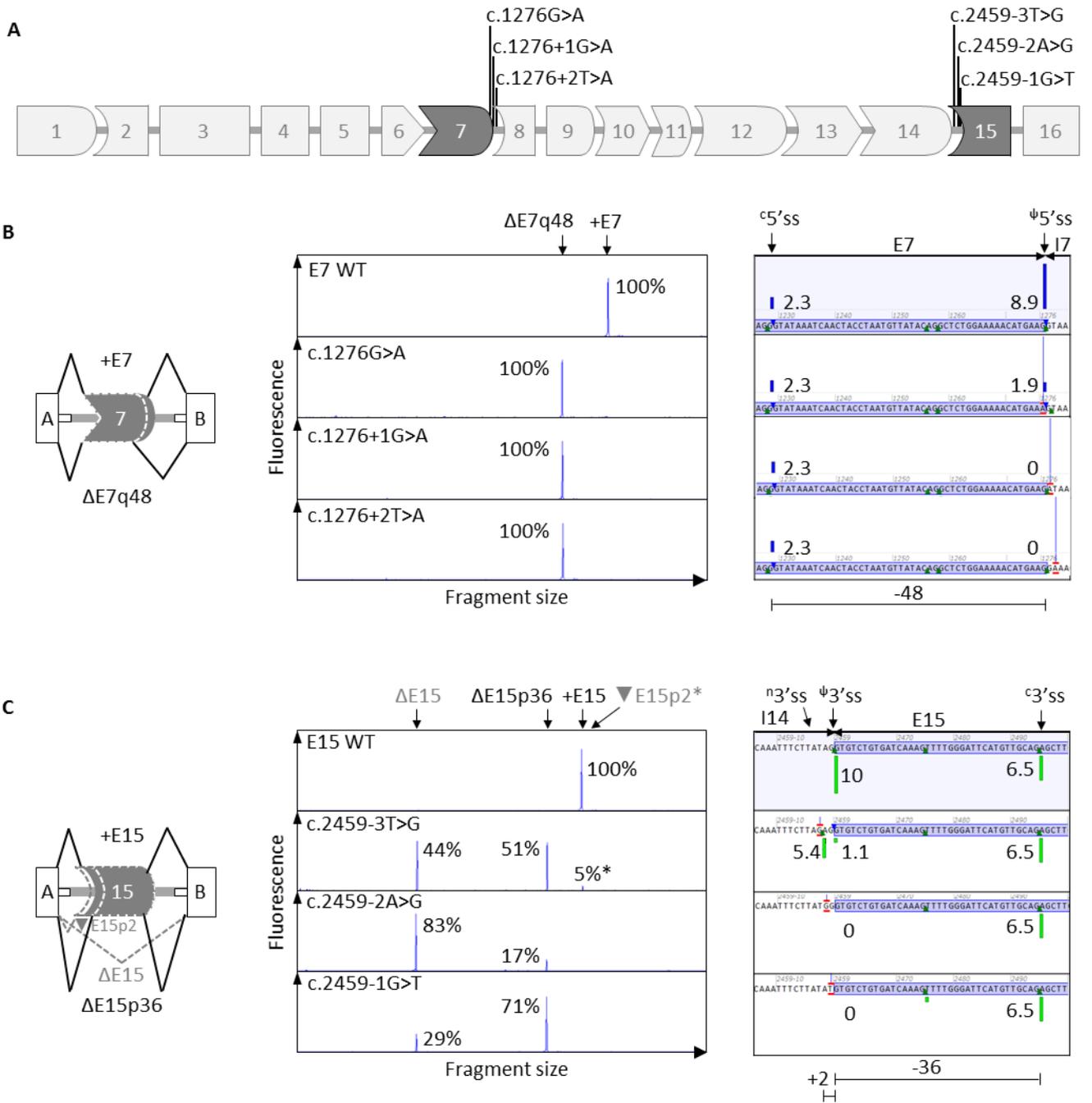


Figure 3

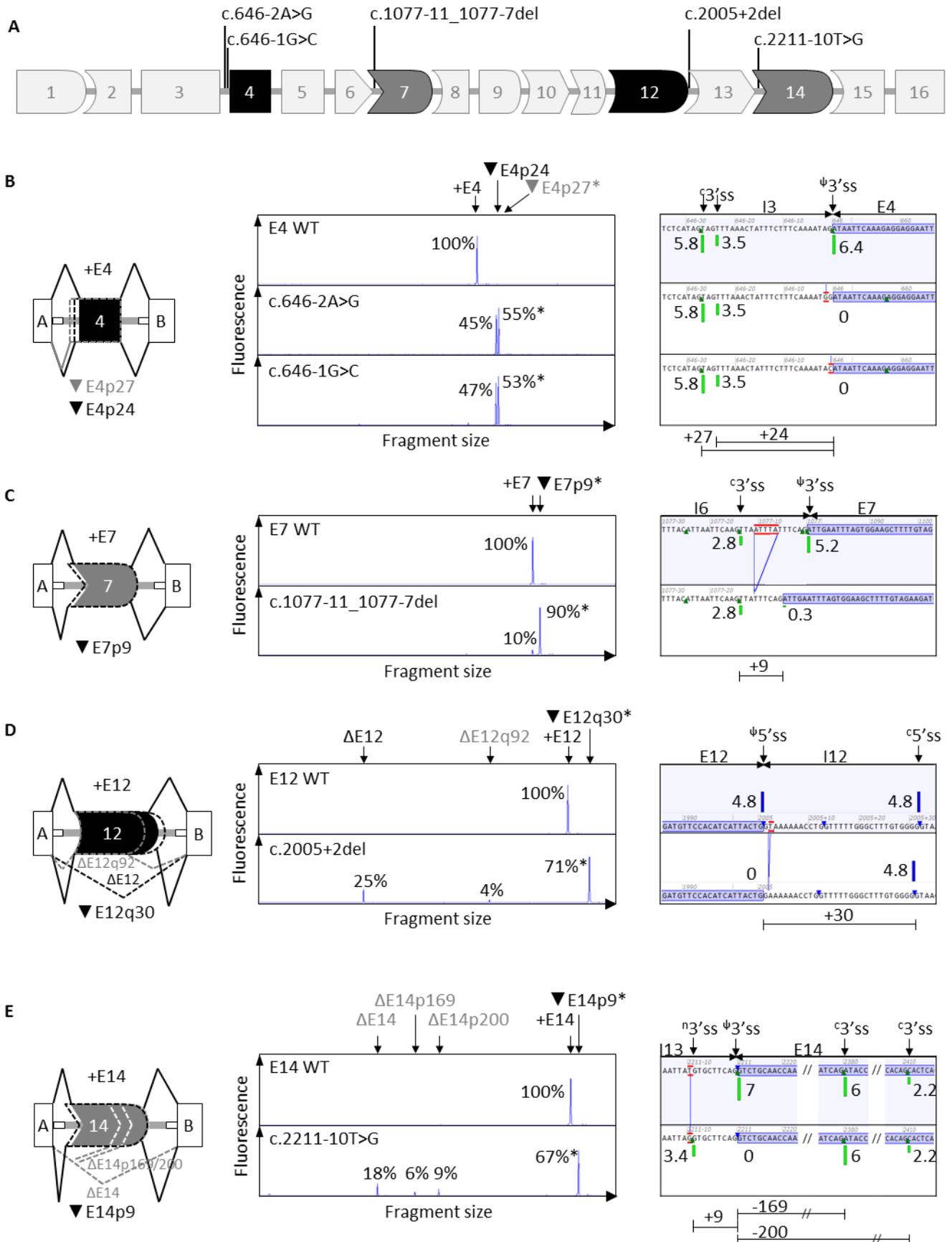


Figure 4

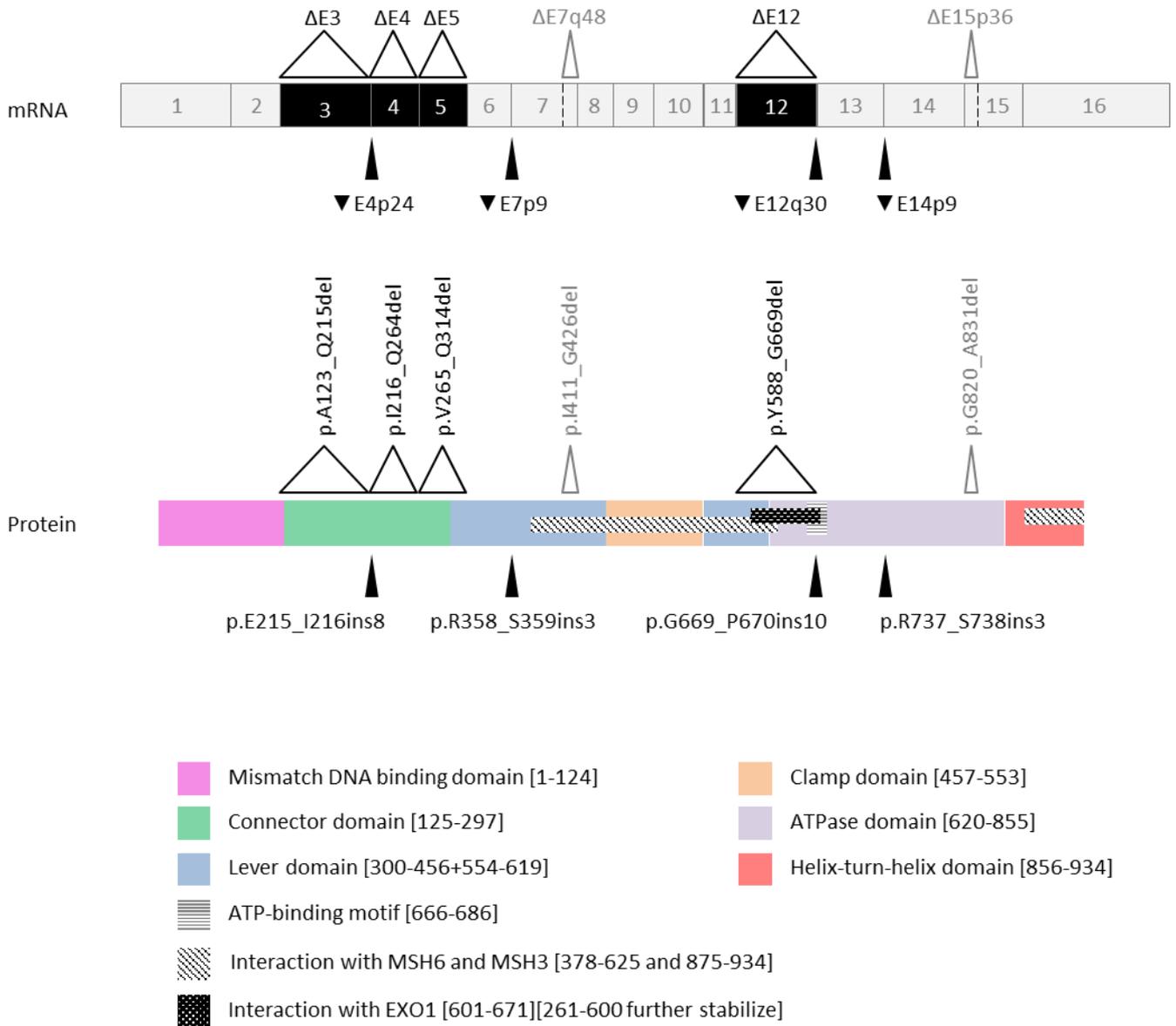
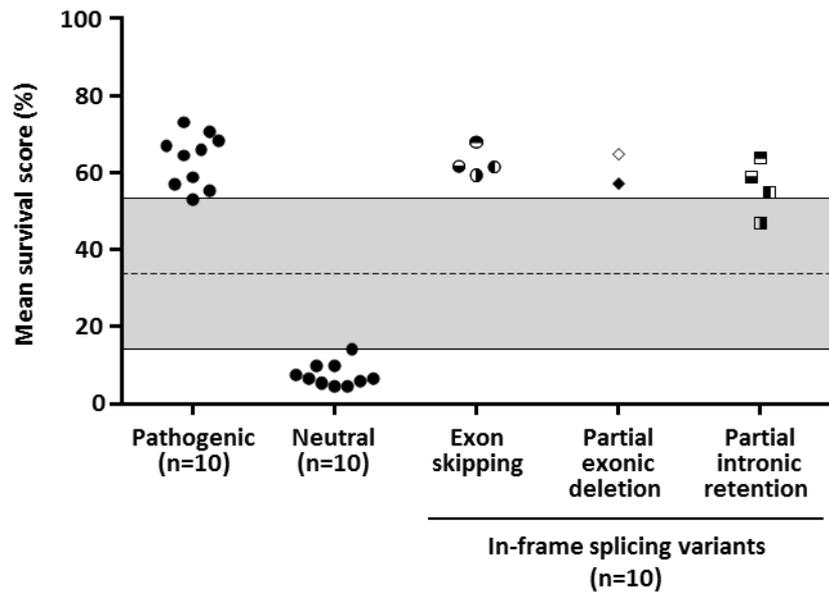


Figure 5



- p.A123_Q215del (ΔE3)
- p.I216_Q264del (ΔE4)
- p.V265_Q314del (ΔE5)
- p.Y588_G669del (ΔE12)
- ◇ p.I411_G426del (ΔE7q48)
- ◆ p.G820_A831del (ΔE15p36)
- p.E215_I216ins8 (▼E4p24)
- p.R358_S359ins3 (▼E7p9)
- p.G669_P670ins10 (▼E12q30)
- p.R737_S738ins3 (▼E14p9)

Supplementary Materials and Methods

Human variation databases

MSH2 variants were retrieved from the following human variation databases (last accessed: 01 January 2020): UMD-*MSH2* (<http://www.umd.be/MSH2/>) [1], ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) [2], HGMD (Human Gene Mutation Database, <http://www.hgmd.cf.ac.uk/ac/index.php>) [3], LOVD (Leiden Open Variation Database, <https://databases.lovd.nl/shared/variants/MSH2/unique>) [4], COSMIC (Catalogue of Somatic Mutations in Cancer, <http://cancer.sanger.ac.uk/cosmic>) [5], dbSNP (Single Nucleotide Polymorphism database, <https://www.ncbi.nlm.nih.gov/projects/SNP/>) [6] and gnomAD (genome Aggregation Database, <http://gnomad.broadinstitute.org/>) [7].

Cell culture

HeLa cell line is derived from human cervical cancer. This cell line was grown in a humidified atmosphere containing 5% CO₂ in Dulbecco's modified Eagle's Medium (ThermoFisher) supplemented with 10% fetal calf serum and was seeded 24 hours before the transfection into 12-well plates at a density of 6.10⁴ cells/well.

LoVo cell line is derived from human colorectal carcinoma and is *MSH2*-deficient [8]. This cell line was grown in a humidified atmosphere containing 5% CO₂ in Dulbecco's modified Eagle's Medium GlutaMAX (ThermoFisher) supplemented with 10% fetal calf serum and was seeded 24 hours before the transfection into 6-well plates at a density of 5.10⁵ cells/well.

Cell-based minigene splicing assay

In order to assess variant-induced impact on splicing, we performed a splicing reporter assay based on the use of the two-exon minigene pCAS2 vector, as previously described [9–11], with minor modifications. Briefly, wild-type (WT) and mutant segment encompassing *MSH2* exon of interest and their intronic flanking sequences were amplified by PCR using specific primers (Supplementary Table S1) from genomic DNA (gDNA) of patients carrying heterozygous variants. The corresponding genomic fragments were then cloned into the intron of the two-exon pCAS2 minigene using *BamHI* and *MluI* restriction sites. When gDNA was not available, the variant of interest was introduced by site-directed mutagenesis (Supplementary Table S1). WT and mutant pCAS2-*MSH2* minigenes were transfected in parallel into HeLa cells using 400 ng of each minigene plasmid and the FuGene 6 Transfection Reagent (Roche Applied Science). Total RNA was extracted 24 hours post-transfection using the NucleoSpin RNA kit (Macherey-Nagel). The splicing patterns were then analyzed by semi-quantitative fluorescence RT-PCR (22 cycles) by using the OneStep RT-PCR kit (Qiagen), in a 25 µL reaction volume, with 200 ng total RNA and specific minigene primers (Supplementary Table S1). RT-PCR products were separated

by electrophoresis on 2% agarose gels. All RT-PCR DNA bands were gel-purified and sequenced. Because of the possible formation of heteroduplexes between the different RT-PCR products, gels image were not used for quantification. For semi-quantitative analysis, fluorescence RT-PCR products were resolved under denaturing conditions by capillary electrophoresis on an automated 3500 Genetic Analyzer (Applied Biosystems). Electropherograms were then analyzed by using the GeneMapper V5.0 Software (Applied Biosystems) and peak areas were used to quantify the relative levels of each transcript.

Methylation tolerance-based functional assay

The methylation tolerance-based assay was performed as previously described [12] with minor changes. The expression construct for the mutant variants comprised a modified version of the pVAX1 expression vector containing the full-length coding sequences of the WT *MSH2* (NM_000251.2) and the green fluorescent protein (*GFP*) reporter genes, separated by an IRES sequence, the expression of the two proteins being under the control of the cytomegalovirus promoter. *MSH2* deletions or insertions were introduced into this vector by site-directed mutagenesis using specific primers (Supplementary Table 1).

Briefly, following transient transfection of LoVo cells with the *MSH2* variant-GFP construct, GFP-positive cells were sorted using a S3 Biorad sorter, seeded into 12-well plates and exposed to a methylating agent, *i.e.* MNNG (N-Methyl-N'-Nitro-N-Nitrosoguanidine) (1 μ M) in 1 or 2 treatments separated by 7 hours. After an incubation period of 11 days post-treatment, methylation tolerance was evaluated using a colony formation assay. The survival fraction was expressed as the percentage of colonies in the treated sample relative to untreated control sample and the mean survival score expresses the average of the survival fractions after 1 and 2 MNNG treatments. Since experiments were conducted in a different lab than that of the original publication, the validation set of 20 previously classified *MSH2* missense variants including 10 known pathogenic and 10 known neutral alterations (*i.e.* all variants of the original publication except the c.593A>G variant, which has been re-classified as class 3 in the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database since publication) was screened as controls in parallel. All samples were tested in triplicate and each experiment was conducted at least in duplicate.

References

- 1 Grandval P, Fabre AJ, Gaildrat P, Baert-Desurmont S, Buisine M-P, Ferrari A, Wang Q, Beroud C, Olschwang S. UMD-MLH1/MSH2/MSH6 databases: description and analysis of genetic variations in French Lynch syndrome families. *Database* 2013;bat036.
- 2 Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016;44:D862–8.
- 3 Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017;136:665–77.
- 4 Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011;32:557–63.
- 5 Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 2017;45:D777–83.
- 6 Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- 7 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, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
- 8 Umar A, Boyer JC, Thomas DC, Nguyen DC, Risinger JI, Boyd J, Ionov Y, Perucho M, Kunkel TA. Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. *J Biol Chem* 1994;269:14367–70.
- 9 Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* 2010;653:249–57.
- 10 Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet* 2016;12:e1005756.
- 11 Meulemans L, Mesman RLS, Caputo SM, Krieger S, Guillaud-Bataille M, Caux-Moncoutier V, Léone M, Boutry-Kryza N, Sokolowska J, Révillion F, Delnatte C, Tubeuf H, Soukarieh O, Bonnet-Dorion F, Guibert V, Bronner M, Bourdon V, Lizard S, Vilquin P, Privat M, Drouet A, Grout C, Calléja FMGR,

Golmard L, Vrieling H, Stoppa-Lyonnet D, Houdayer C, Frebourg T, Vreeswijk MPG, Martins A, Gaildrat P. Skipping Nonsense to Maintain Function: The Paradigm of *BRCA2* Exon 12. *Cancer Res* 2020;80:1374–86.

12 Bouvet D, Bodo S, Munier A, Guillerm E, Bertrand R, Colas C, Duval A, Coulet F, Muleris M. Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the *MLH1* and *MSH2* Genes and Identify Patients With Lynch Syndrome. *Gastroenterology* 2019;157:421–31.

Figure S1

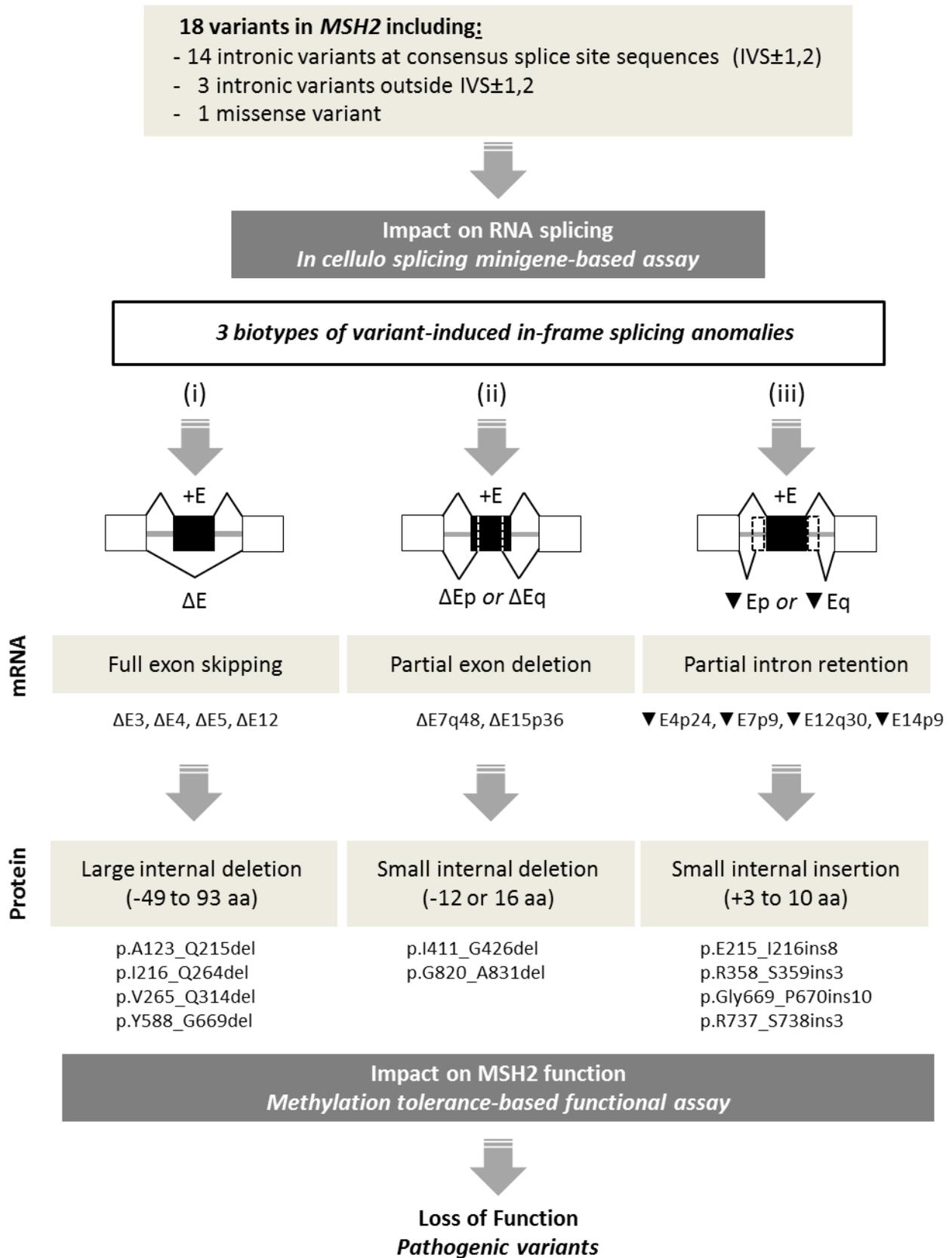


Table S1. Description of the primers used in this study

Primer ¹	Location	Sequence ²	Use
MSH2E3-BamHI-F	MSH2	I2 GACCGGATCCTCTAGCCTGGGTGACAGAGC	Genomic DNA PCR amplification for minigene cloning
MSH2E3-MluI-R		I3 GACCACGCGTACAAAGTTTCTGTAAC	
MSH2E4-IFBamHI-F		I3 AGAAGTGCAGGATCCTGCACATTGTATAAACATTTAATGT	
MSH2E4-IFMluI-R		I4 CAAAACAAGACGCGTGAGATAAATATGACAGAAATATC	
MSH2E5-BgIII-F		I4 GACCAGATCTTTTGGCTATTCTAAATAATGCTG	
MSH2E5-MluI-R		I5 GACCACGCGTAAAAAGTGGAGTGGAGGAGGG	
MSH2E7-BamHI-F1 ⁴		I6 GACCAGGATCCTACTGCGCCAGCAGAT	
MSH2E7-MluI-R1 ⁴		I7 GACCACGCGTGGCAACGGAGCAAGAC	
MSH2E7-BamHI-F2 ⁴		I6 GACCAGGATCCTGCCCAGCAGATTCAGCTTT	
MSH2E7-MluI-R2 ⁴		I7 GACCACGCGTCTGAATGTGTCTAAGAGTGAG	
MSH2E12-IFBamHI-F		I11 AGAAGTGCAGGATCCCTATGTGTAGTTTTAGGTGGGTTCT	
MSH2E12-IFBamHI-R		I12 CAAAACAAGACGCGTTTTCTAATGTTAAGAAGTGGGAATT	
MSH2E14-BamHI-F		I13 GACCGGATCCTCATTATGTTTTGGCATATCCT	
MSH2E14-MluI-R		I14 GACCACGCGTGTGAATTAAGGGGTAATGGG	
MSH2E15-BamHI-F		I14 GACCAGGATCCTGCTAATGTATTTGAAGTAATCCGA	
MSH2E15-MluI-R	I15 GACCACGCGTCACAGAGATAGATTCCTTTGCCAT		
MSH2E3c.645+1GA-F	MSH2	E3/I3 AACTGAGACAGATAAGCAAATTG	Minigene pCAS2-MSH2
MSH2E3c.645+1GA-R		E3/I3 CAATTTGCTTATCTGTCTCAGTT	
MSH2E4c.646-2AG-F		I3/E4 TTTCAAAATGGATAAATCAAAGAG	
MSH2E4c.646-2AG-R		I3/E4 CTCTTTGAATTATCCATTTTGAAA	
MSH2E4c.646-1GC-F		I3/E4 TTTCAAAATACATAAATCAAAGAG	
MSH2E4c.646-1GT-R		I3/E4 CTCTTTGAATTATATATTTTGAAA	
MSH2E4c.792+1GC-F		E4/I4 GGAGAATCAGCTACATGGATTA	
MSH2E4c.792+1GA-R		E4/I4 ATAATCCATGTATCTGATTCTCC	
MSH2E7c.1276+1GA-F		E7/I7 AAAAAATGAAGATAACAAGTG	
MSH2E7c.1276+2TA-R		E7/I7 CACTTGTTCCTTCATGTTTT	
MSH2E12c.1760-2AG-F		I11/E12 TTTTATACGGGCTATGTAGAAC	
MSH2E12c.1760-2AG-R		I11/E12 GTTCTACATAGCCCGTATAAAA	
MSH2E12c.2005+1GA-F		E12/I12 CATCATTACTGATAAAAAACCT	
MSH2E12c.2005+1GA-R		E12/I12 AGGTTTTTTATCAGTAATGATG	
MSH2E12c.2005+2TC-F		E12/I12 CATCATTACTGGCAAAAAACCT	
MSH2E12c.2005+2TC-R	E12/I12 AGGTTTTTTGCCAGTAATGATG		
MSH2E15c.2459-2AG-R	I14/E15 ATCACAGACACCATAAGAAATTT		
MSH2E15c.2459-1GT-F	I14/E15 AAATTTCTTATATGTGCTGTGAT		
pCASseq-F	pCAS2 minigene	IA GGGGTCAATAGCAGTGAGAG	Site directed mutagenesis pCAS2 (PCR1 and PCRII) and sequencing
pCASseq-R		IB GCTCCATTTCACAGGTAGAGA	
pCASKO1-F*	pCAS2 minigene	EA TGACGTGCGCCCGCCATCAC	Analysis of MSH2 splicing patterns by RT-PCR and sequencing
pCAS2-R		EB ATTGGTTGTTGAGTTGGTTGTC	
pVAX1MS2E1-XmaI-F	MSH2	E1/E2 TGCTGGCCGCCCGGGAGGTGTTCAAGACCCAGGG	Site directed mutagenesis pVAX1
pVAX1MS2E13-MfeI-F		E13 GGTGACAGTCAATTGAAAGGAGTCTCCACGTTTCATG	
pVAX1MS2E13-MfeI-R		E13 GACTCCTTTCAATTGACTGTCCACCAGCCCTACTCG	
pVAX1MS2E13-BamHI-R		E16/ires GAGAGGGGCGGATCCTCACGTAGTAACCTTTTATTCTG	
pVAX1-delE3-F		E2/E4 TATTTGGCATATAAGATAAATCAAAGAGGAGGAATTC	
pVAX1-delE3-R		E2/E4 TCCTCTTTGAATTATCTTATATGCCAAATACCAATCA	
pVAX1-delE4-F		E3/E5 GGGAAACTGAGACAGGTTGCAGTTTCATCACTGTCT	
pVAX1-delE4-R		E3/E5 TGATGAACTGCAACCTGTCTCAGTTTCCCCATGT	
pVAX1-delE5-F		E4/E6 GAAATGGAGAACTCAGGGTTCTGTTGAAGATACCACT	
pVAX1-delE5-R		E4/E6 ATCTTCAACAGAACCTGATTCTCCATTTCTGGCA	
pVAX1-delE12-F		E11/E13 TCAATATTTCTCAGGCCCAATATGGGAGGTA	
pVAX1-delE12-R		E11/E13 CTCCCATATTTGGGGCTGAAGAAATATTGACAATTTCTT	
pVAX1-insE4p24-F		E4 TTTAAACTATTTCTTCAAATGGATAAATCAAAGAGGAGGAATCT	
pVAX1-insE4p24-R		E4 CCATTTTGAAGAAATAGTTTAAACTGTCTCAGTTTCCCCATGT	
pVAX1-insE7p9-F		E6/E7 AGAGGAGAGTTATTTAGATTGAATT	
pVAX1-insE7p9-R		E6/E7 AATTCATCTGAAATAAATCTCTCTCT	
pVAX1-delE7q48-F		E7/E8 ACCGACTCTATCAGGGAAAACACCAGAAATATTGTTG	
pVAX1-delE7q48-R		E7/E8 ATTTCTGGTGTTCCTCTGATAGAGTCGGTAACAA	
pVAX1-insE12q30-F		E12 GAAAAAACCTGGTTTTGGGCTTTGTGGGGCCCAATATGGGAGGTAA	
pVAX1-insE12q30-R		E12 CCCACAAAAGCCCAAAACAGGTTTTTCCAGTAATGATGTGGAACATC	
pVAX1-insE14p9-F	E13/14 TATCCTCAGGTGCTTCAGGTCTGCAACC		
pVAX1-insE14p9-R	E13/14 GGTTGCAGACCTGAAGCACCTGAGGATA		
pVAX1-delE15p36-F	E14/E15 ATCAGGTGAAGAAGAGCTTGTCAATTTCCCTAAGC		
pVAX1-delE15p36-R	E14/E15 GGAATTAGCAAGCTCTTTCTTACCTGATAAAGCAT		
MSH2E6-R	E6 TCTTTGCTCTGAGGGTTTT	Sequencing pVAX1	
MSH2E9-R	E9 CAAGATCTCTGGCTGCACTTA		
MSH2E8-F	E8 AGAAACAACCTTTAGATATGGATC		
MSH2E11-R	E11 GAAGAAGCCAGGATGCC		
MSH2E11-F	E11 ATCCTGGGCTTCTCATATTC		
pVAX1-pCMV-F	cmv AGGTCTATATAAGCAGAGCTCT		
pVAX1-ires-R	iresR GCAATATGGTGGAAAATAACAT		

¹ F, forward; R, reverse; *, fluorescent primer

² Intronic and exonic sequences are indicated in grey and black, respectively. The position of the variant is underlined>. The double underlined sequences correspond to restriction sites for BamHI, BglII, MluI, XmaI and MfeI.

³ The forward primers listed were used with the reverse primer pCAS-seq-R and the reverse primers listed were used with the forward primer pCAS-seq-F for mutagenesis.

⁴ The MSH2E7-BamHI-F1/MluI-R1 primer set was used for the amplification of the genomic DNA carrying the variants c.1276G>A, 1276+1G>A, c.1276+2T>A whereas the MSH2E7-BamHI-F2/MluI-R2 primer set was used for the amplification of the genomic DNA carrying the variant c.1077-11_1077-7del.

Table S2. MSH2 variants selected in this study from different variation databases

Selection criteria: In-frame splicing alterations	Variant	Predicted amino acid change	Location	Databases ¹						
				UMD-MSH2	ClinVar	HGMD	LOVD / InSiGHT	COSMIC	dbSNP	gnomAD
Exon skipping n = 7	c.645+1G>A	p.?	I3	5	5	DM	5	v	5	
	c.792+1G>A	p.?	I4		5	DM	5	v	5	
	c.792+1G>C	p.?	I4	5						
	c.942+1_942+2del	p.?	I5	5						
	c.1760-2A>G	p.?	I11				4	v		
	c.2005+1G>A	p.?	I12	5	5/4	DM	5/4	v	5/4	
	c.2005+2T>C	p.?	I12		5	DM	5		5	
Partial exon deletion n = 6	c.1276G>A	p.G426R	E7		3	DM	4		3	
	c.1276+1G>A	p.?	I7	5	5	DM	5		5/4	
	c.1276+2T>A	p.?	I7	5	4	DM	4		4	
	c.2459-3T>G	p.?	I14	4	3					
	c.2459-2A>G	p.?	I14		5	DM	4		5	
	c.2459-1G>T	p.?	I14	5						
Partial intron retention n = 5	c.646-2A>G	p.?	I4	5	4	DM	5/4		4	
	c.646-1G>C	p.?	I4	5	5		5		5	
	c.1077-11_1077-7del	p.?	I6	3						
	c.2005+2del	p.?	I12	5	4	DM	5/4		4	
	c.2211-10T>G	p.?	I13	3			5			

¹ Variant classification as reported in UMD-MSH2, ClinVar, LOVD / InSiGHT and dbSNP are indicated: 3, uncertain significance; 4, likely pathogenic; 5, pathogenic. For HGMD, DM = disease causing mutation and for COSMIC, v = variant reported without classification.

Table S3. Characterization of *MSH2* spliceogenic in-frame variants in splicing minigene-based assays

In-frame splicing alterations	Variant	Location	Impact on splicing: transcript levels (%)			
			Physio	In-Frame	Frame-shift	
Exon skipping			+E3	$\Delta E3$		
	WT E3		100 ± 0	0 ± 0		
	c.645+1G>A	I3	0 ± 0	100 ± 0		
			+E4	$\Delta E4$	▼ E4q38	
	WT E4		100 ± 0	0 ± 0	0 ± 0	
	c.792+1G>A	I4	0 ± 0	93 ± 0.6	7 ± 0.6	
	c.792+1G>C	I4	0 ± 0	83 ± 3.2	17 ± 3.2	
			+E5	$\Delta E5$		
	WT E5		100 ± 0	0 ± 0		
	c.942+1 +2del	I5	0 ± 0	100 ± 0		
			+E12	$\Delta E12$	$\Delta E12q92$	▼ E12q31
	WT E12		100 ± 0	0 ± 0	0 ± 0	0 ± 0
	c.1760-2A>G	I11	0 ± 0	100 ± 0	0 ± 0	0 ± 0
	c.2005+1G>A	I12	0 ± 0	53 ± 3.8	13 ± 0.6	34 ± 3.2
c.2005+2T>C	0 ± 0		43 ± 1.3	6 ± 0	51 ± 1.3	
Partial exon deletion			+E7	$\Delta E7q48$		
	WT E7		100 ± 0	0 ± 0		
	c.1276G>A	E7	0 ± 0	100 ± 0		
	c.1276+1G>A	I7	0 ± 0	100 ± 0		
	c.1276+2T>A	I7	0 ± 0	100 ± 0		
			+E15	$\Delta E15p36$	▼ E15p2	$\Delta E15$
	WT E15		100 ± 0	0 ± 0	0 ± 0	0 ± 0
	c.2459-3T>G	I14	0 ± 0	51 ± 5.2	5 ± 0.8	44 ± 5.9
	c.2459-2A>G		0 ± 0	17 ± 2.6	0 ± 0	83 ± 2.6
c.2459-1G>T	0 ± 0		71 ± 10.2	0 ± 0	29 ± 10.2	
Partial intron retention			+E4	▼ E4p24	▼ E4p27	
	WT E4		100 ± 0	0 ± 0	0 ± 0	
	c.646-2A>G	I3	0 ± 0	45 ± 1	55 ± 1	
	c.646-1G>C	I3	0 ± 0	47 ± 0	53 ± 0	
			+E5	$\Delta E5$		
	WT E7		100 ± 0	0 ± 0		
	c.1077-11_1077-7del	I6	10 ± 1	90 ± 1		
			+E12	$\Delta E12$	▼ E12q30	$\Delta E12q92$
	WT E12		100 ± 0	0 ± 0	0 ± 0	0 ± 0
	c.2005+2del	I12	0 ± 0	25 ± 5.2	71 ± 5.1	4 ± 0.6
			+E14	▼ E14p9	$\Delta E14p169$	$\Delta E14p200$
WT E14		100 ± 0	0 ± 0	0	0	0
c.2211-10T>G	I13	0 ± 0	67 ± 9.2	9 ± 1	6 ± 1	18 ± 7.4

Table S4. MSH2 transcript and protein isoforms resulting from variant-induced splicing anomalies

Splicing anomalies		Transcript isoforms		Alternative ¹	+/- nt	Protein isoforms	+/- aa	Functional domains affected
In-frame	Exon skipping	ΔE3	r.367_645del	y	-279	p.(A123_Q215del)	-93	Mismatch binding domain / Connector domain
		ΔE4	r.646_792del	n	-147	p.(I216_Q264del)	-49	Connector domain
		ΔE5	r.793_942del	y	-150	p.(V265_Q314del)	-50	Connector domain / Lever domain
		ΔE12	r.1760_2005	n	-246	p.(Y588_G669del)	-82	Lever domain / Interaction with MSH6 / Interaction with EXO1 / ATP-binding motif
	Partial exon deletion	ΔE7q48	r.1229_1276del	n	-48	p.(I411_G426del)	-16	Lever domain / Interaction with MSH6
		ΔE15p36	r.2459_2494del	n	-36	p.(G820_A831del)	-12	ATPase domain
	Partial intron retention	▼E4p24	r.645_646ins24 (r.645_646ins646-24_646-1; 646-2A>G)	n	+24	p.(E215_I216ins8), p.(E215_I216insFKLFLSKW)	+8	Connector domain
		▼E7p9	r.1076_1077ins9 (r.1076_1077ins1077-14_1077-1; 1077-11_1077-7del)	n	+9	p.(R358_S359ins3), p.(R358_S359insLFQ)	+3	Lever domain
		▼E12q30	r.2005_2006ins30 (r.2005_2006ins2005+1_2005+31; 2005+2del)	n	+30	p.(G669_P670ins10), p.(G669_P670insEKTWFLGFGV)	+10	Interaction with EXO1 / ATP-binding motif
		▼E14p9	r.2210_2211ins9 (r.2210_2211ins2211-9_2211-1)	n	+9	p.(R737_S738ins3), p.(R737_S738insVLQ)	+3	ATPase domain
Frame-shift or Stop gain	Exon skipping	ΔE14	r.2211_2458del	n	-248	p.(S738Cfs*3)		
		ΔE15	r.2459_2634del	n	-176	p.(G820Afs*3)		
	Partial exon deletion	ΔE14p170	r.2211_2379del	n	-170	p.(S738Yfs*18)		
		ΔE14p201	r.2211_2410del	n	-201	p.(R737Sfs*5)		
		ΔE12q92	r.1914_2005del	n	-92	p.(H639Pfs*6)		
	Partial intron retention	▼E4p27	r.645_646ins27 (r.645_646ins646-27_646-1, 646-2A>G)	n	+27	p.(I216*)		
		▼E4q38	r.792_793ins38 (r.792_793ins792+1_792+38; 792+1G>C)	n	+38	p.(V265Lfs*3)		
		▼E12q31	r.2005_2006ins31 (r.2005_2006ins2005+1_2005+31; 2005+2T>C)	n	+31	p.(P670Kfs*16)		
		▼E15p2	r.2458_2459ins2 (r.2458_2459ins2459-2_2459-1)	n	+2	p.(G820Efs*22)		

¹ Alternative transcripts (naturally occurring RNA), as described in Thompson et al, 2015 (y, yes; n, no)

Table S5. Patient clinical, tumoral and family data

In-frame splicing alterations	MSH2 variant	Patient	Patient tumor*, ** (age)	MSI status	Immunohistochemistry (IHC)				Tumor* in relatives*** (age)
					MLH1	MSH2	MSH6	PMS2	
Exon skipping	c.645+1G>A	P1	OC (46)	ND	ND	ND	ND	ND	ND
		P2	CRC (58)	MSI	present	present	loss	present	ND
		P3	ND	ND	ND	ND	ND	ND	ND
	c.792+1G>A	No ⁰	ND	ND	ND	ND	ND	ND	ND
	c.792+1G>C	P4	ND	ND	ND	ND	ND	ND	ND
	c.942+1_942+2del	P5	ND	ND	ND	ND	ND	ND	ND
	c.1760-2A>G	P6	KC (74), EC** (75), CRC (77)	MSI	present	present	loss	present	C: CRC
	c.2005+1G>A	P7	CRC (39), SC** (57)	ND	present	loss	loss	present	F: HNC (62,66); M: CC (64)
		P8	EC (51), CRC (65)	ND	ND	ND	ND	ND	M: CRC (42); So: CRC (35); D: EC and OC (42)
P9		CRC (26)	MSI	present	loss	loss	present	ND	
c.2005+2T>C	No ⁰	ND	ND	ND	ND	ND	ND	ND	
Partial exon deletion	c.1276G>A	P10	ND	ND	ND	ND	ND	ND	ND
	c.1276+1G>A	P11	ND	MSI	present	loss	loss	present	ND
	c.1276+2T>A	P12	ND	MSI	present	loss	loss	present	ND
		P13	ND	ND	ND	ND	ND	ND	ND
		P14	ND	ND	ND	ND	ND	ND	ND
		P15	ND	ND	ND	ND	ND	ND	ND
	c.2459-3T>G	P16	CRC (42), SC (43), LC (45-50), PC (50), BC** (66)	MSI	present	present	weak	present	ND
	c.2459-2A>G	P17	CRC (44)	MSI	present	loss	loss	present	ND
		P18	CRC (38)	MSI	present	loss	loss	present	ND
c.2459-1G>T	P19	EC (42), CRC (44)	ND	ND	ND	ND	ND	M: CRC (50,70), EC (50)	
Partial intron retention	c.646-2A>G	P20	EC (40)	ND	ND	ND	ND	ND	U: CRC; A: EC
		P21	CRC (49)	MSS	present	loss	loss	ND	ND
		P22	CRC (40,63)	MSI	present	loss	loss	ND	Si: EC (42); B: CRC (45); M: EC (47); U: CRC (50); A: CRC (35)
		P23	ND	ND	ND	ND	ND	ND	ND
	c.646-1G>C	P24	CRC (64)	MSI	loss	loss	loss	ND	ND
		P25	ND	ND	ND	ND	ND	ND	ND
	c.1077-11_1077-7del	P26	CRC (<50)	MSI	present	loss	loss	present	ND
	c.2005+2del	P27	ND	MSI	present	loss	loss	present	ND
		P28	ND	ND	ND	ND	ND	ND	ND
		P29	EC (43)	ND	ND	ND	ND	ND	M: EC?/CRC?/BrC? (38); Si: CRC (38); B: KC (52)
		P30	EC (40)	ND	ND	ND	ND	ND	F: CRC (48); U: CRC? (54) ; A: EC (57)
		P31	CRC (29), G** (52)	ND	weak	weak	ND	weak	F: StC (49); M: EC (63) ; B: 2 CRC (20,39) ; Si: EC (43) ; So: S (19)
		P32	ND	ND	ND	ND	ND	ND	ND
		P33	ND	ND	ND	ND	ND	ND	ND
	c.2211-10T>G	P34	EC	MSI	present	loss	loss	present	GM: CRC (63); C: CRC (54), C: BrC (58); C: CRC (40)
		P35	CRC (53)	MSI	present	loss	loss	present	ND
P36		CRC** (59), BrC (52), S (61), LC (70)	MSI	present	present	ND	present	B: PC (51); U: PC (51); N: BrC (47); N: CRC; M: BrC (60); A: BrC (81)	
P37		CRC** (47), EC (61)	MSI	present	present	ND	ND	ND	
P38		CRC	MSI	present	present	loss	ND	ND	
P39		ND	ND	ND	ND	ND	ND	ND	
P40		CRC (48)	MSS	ND	ND	ND	ND	ND	

* CRC, colorectal cancer; EC, endometrium cancer; KC, kidney cancer; SC, small intestine cancer; PC, prostate cancer; BC, urothelial carcinoma of the bladder cancer; LC, lung cancer; HNC, head and neck cancer; CC, cervix cancer; OC, ovarian cancer; BrC, breast cancer; S, sarcoma; G, glioblastoma; StC, stomach cancer

** Tumor analyzed for evaluation of MSI status and protein expression (IHC)

*** M, mother; F, father; D, daughter; So, son; Si, sister; B, brother; A, aunt; U, uncle; GM, grandmother; C, cousin; N, niece or nephew

No^o, no patient reported in the French database; ND, no data

Table S6. Comparison of the results obtained in minigene-based assays in this study with previously reported data generated from patient RNA analyzes: identification of variant-induced in-frame splicing RNA isoforms

<i>MSH2</i> variant	Minigene data (this study)	Patient RNA data (previously published and/or reported in databases)
c.645+1G>A	ΔE3 (100%)	LCL: ΔE3 (major transcript) and ΔE3q154 (minor transcript) (Thompson <i>et al</i> , 2013) Patient RNA: ΔE3 (D. du Sart, LOVD/Insight)
c.792+1G>A	ΔE4 (93%) , ▼E4q38 (7%)	Diploid-haploid conversion assay: ΔE4 (Casey <i>et al</i> , 2005)
c.1276G>A	ΔE7q48 (100%)	LCL: ΔE7q48 (Vargas-Parra <i>et al</i> , 2017) (Morak <i>et al</i> , <i>in preparation</i>)
c.1276+1G>A	ΔE7q48 (100%)	Paxgene: ΔE7q48 (Betz <i>et al</i> , 2010)
c.1276+2T>A	ΔE7q48 (100%)	LCL: ΔE7q48 (major transcript) and ΔE7 (minor transcript) (Auclair <i>et al</i> , 2006)
c.2005+2del	▼ E12q30 (71%) , ΔE12 (25%), ΔE12q92 (4%)	Paxgene: ▼ E12q30 (unpublished data from our group reported in the French UMD-MSH2 database, as mentioned in Thompson <i>et al</i> , 2020)
c.2211-10T>G	▼ E14p9 (67%) , ΔE14 (18%), ΔE14p200 (9%), ΔE14p169 (6%)	LCL: ▼ E14p9 (unpublished data from our group reported in the French UMD-MSH2 database, as mentioned in Thompson <i>et al</i> , 2020)

In **bold**, in-frame splicing RNA isoforms

LCL, Lymphoblastoid cell line

Discussion

Mes travaux de thèse s'inscrivent dans la problématique de l'interprétation biologique et clinique des variations de séquence. À l'ère de la médecine génomique de précision, l'évaluation de la pathogénicité des variations génétiques identifiées dans le génome des patients représente un des défis majeurs dans le domaine de la génétique médicale. L'interprétation clinique des changements de séquence au niveau des gènes impliqués dans les maladies monogéniques repose généralement sur un faisceau d'arguments prenant en compte, notamment : (i) les données phénotypiques des individus porteurs des variations, (ii) la fréquence allélique des variations dans la population générale et dans les cohortes de patients, (iii) la co-ségrégation des variations avec la maladie, (iv) la co-occurrence des variations en *cis* ou en *trans* dans le même gène, (v) les analyses somatiques, (vi) des conséquences biologiques supposées de la variation (nature de la variation et prédictions bio-informatiques) ainsi que (vii) des données expérimentales fonctionnelles. En 2015, le consortium ACMG-AMP a proposé des recommandations qui établissent un cadre précis pour l'interprétation des variations en attribuant à chaque argument une direction en faveur de la pathogénicité ou de la neutralité mais également une pondération qui rend compte de la force de chaque évidence (Richards *et al.*, 2015). Le seul critère de pathogénicité défini comme très fort (PVS1, *Pathogenic Very Strong*) est attribué aux variations présumées nulles, telles que les variations non-sens, *frameshift* ou $IVS\pm 1/2$, responsables *a priori* d'une perte de fonction du fait de l'introduction d'un PTC. Les données acquises dans le cadre de mes travaux de thèse ont permis de valider ou au contraire d'invalider le fait que certaines de ces variations sont bien responsables d'une perte totale de fonction, avec des conséquences importantes pour l'interprétation clinique de ces variations. Notre démonstration s'est appuyée essentiellement sur l'évaluation de l'impact de ces variations sur l'épissage de l'ARN.

Il est maintenant clairement établi qu'une fraction importante des variations auparavant considérées comme étant de signification inconnue (variations synonymes, faux-sens ou introniques en dehors de $IVS\pm 1/2$) est pathogène du fait de leurs effets sur l'épissage, ces altérations conduisant généralement à la perte de fonction du fait d'un décalage du cadre de lecture avec l'introduction d'un PTC (Rebbek *et al.*, 2018 ; Thompson *et al.*, 2020). Notre équipe de recherche a acquis une expertise significative dans le développement et l'implémentation de tests fonctionnels axés sur la mise en évidence expérimentale des défauts d'épissage de l'ARN et a ainsi contribué à la classification neutre ou pathogène de nombreuses VSI splicéogéniques dans les gènes *MMR* et *BRCA* (Tournier *et al.*, 2008 ; Théry *et al.*, 2011 ; Gaildrat *et al.*, 2012 ; Houdayer *et al.*, 2012 ; Di Giacomo *et al.*, 2013 ; Soukarieh *et al.*, 2016 ; Tubeuf *et al.*, 2020a). Dans le cadre de mes travaux de thèse, c'est un changement de paradigme qui a été envisagé en posant cette fois l'hypothèse selon laquelle certaines variations supposées pathogènes pouvaient maintenir une fonction, au moins en partie, du fait de leur impact sur l'épissage à l'origine de modifications en phase. Nous nous sommes principalement intéressées

aux variations supposées nulles de type IVS \pm 1/2 et non-sens, localisées dans 2 gènes majeurs de prédisposition aux cancers, *BRCA2* pour le syndrome sein-ovaire et *MSH2* pour le syndrome de Lynch.

I. Les variations présumées pathogènes sont susceptibles de maintenir une fonction par le biais de leur impact sur l'épissage

I.1. Les variations IVS \pm 1/2 responsables d'un défaut d'épissage en phase et leurs conséquences fonctionnelles au niveau de la protéine

I.1.1. Les variations IVS \pm 1/2 à l'origine d'un saut d'exon en phase

Dans le cadre de mes travaux de thèse, nous avons sélectionné, à partir de bases de données mutationnelles, des variations IVS \pm 1/2 dans les gènes *BRCA2* et *MSH2* susceptibles d'induire un saut d'exon en phase, du fait de la destruction des sites constitutifs 5' ou 3' d'épissage. Nos investigations se sont ainsi focalisées sur les altérations de l'épissage, d'une part, de l'exon 12 de *BRCA2* et, d'autre part, des exons 3, 4, 5 et 12 de *MSH2*. Comme attendu, les analyses expérimentales basées sur l'utilisation de minigènes et celles conduites à partir de l'ARN des patients, lorsque ce matériel était disponible, nous ont permis de montrer que ces variations étaient responsables d'un effet drastique sur l'épissage, c'est-à-dire que dans le contexte muté, le transcrit de référence (pleine longueur) n'est plus produit. Elles sont en effet toutes responsables de la production de transcrits dépourvus d'un exon en phase. Il est toutefois important de noter que certaines d'entre elles génèrent en plus d'autres isoformes aberrantes hors phase. C'est par exemple le cas de la variation *BRCA2* c.6937+2del qui est responsable de la production de deux transcrits aberrants dans des proportions équivalentes : l'un avec le saut en phase de l'exon 12 et l'autre avec la rétention hors phase des 40 premiers nucléotides de l'intron 12, résultat de l'utilisation d'un 5'ss cryptique intronique. De façon surprenante, la destruction du même 5'ss constitutif par deux autres variations, c.6937+1G>A et c.6937+1G>T, ne conduit qu'au saut en phase de l'exon 12. Afin d'expliquer ces différences, nous proposons l'hypothèse selon laquelle la variation c.6937+2del est responsable d'un effet combiné sur le 5'ss mais également sur un élément régulateur de l'épissage chevauchant le 5'ss. D'ailleurs, l'équipe du Dr Spurdle a récemment montré que l'outil HEXplorer prédisait l'altération d'un élément ESS uniquement dans le contexte muté c.6937+2del (Canson *et al.*, 2020). De la même façon, les variations *MSH2* c.2459-2A>G et c.2459-1G>T conduisent toutes les deux à la destruction du 3'ss constitutif et présentent pourtant des effets sur l'épissage drastiquement différents. Alors que la variation c.2459-2A>G conduit à la production majoritaire (83%) du saut hors phase de l'exon 15, la variation c.2459-1G>T produit principalement (71%) une isoforme en phase déléetée pour les 36 premiers nucléotides de l'exon 15, résultat de l'utilisation d'un site 3'ss cryptique exonique. Ici encore, il est envisageable que ces effets

ne soient pas seulement médiés par l'altération du 3'ss mais également par la modification d'un élément régulateur de l'épissage chevauchant le 3'ss constitutif. Afin d'anticiper ce type d'effets concomitants, il serait important de combiner les prédictions bio-informatiques dédiées aux sites d'épissage, telles que celles produites par MaxEntScan, avec celles prenant en compte d'autres déterminants de l'épissage, tel que l'algorithme récemment développé SpliceAI (Jaganathan et al, 2019), SPANR (Xiong *et al.*, 2015) ou d'autres outils spécifiquement dédiés à la détection des SRE (Tubeuf *et al.*, 2020b). Ces données soulignent l'importance (i) de tester expérimentalement l'effet sur l'épissage de toutes les variations, y compris les IVS \pm 1/2, en particulier si un site d'épissage cryptique existe dans le voisinage, et (ii) d'appréhender de façon qualitative et quantitative les conséquences des anomalies d'épissage au niveau des transcrits produits dans le contexte muté.

L'existence d'anomalies d'épissage en phase soulève la question de la fonctionnalité de telles isoformes théoriquement à l'origine de protéines présentant des délétions internes. Le maintien ou l'abrogation de l'activité protéique dépend alors de l'importance fonctionnelle de la séquence délétée. Pour les variations IVS \pm 1/2 responsables du saut en phase de l'exon 12 de *BRCA2*, nous avons montré, dans un test fonctionnel basé sur l'utilisation de cellules souches embryonnaires de souris, qu'elles permettent la survie cellulaire (test de complémentation) et la production d'une protéine fonctionnelle, au moins en partie (activité de recombinaison homologue). Ces données sont en accord avec la démonstration antérieure de la fonctionnalité de la protéine *BRCA2* délétée pour l'exon 12 (Li *et al.*, 2009). Par contraste, les isoformes protéiques délétées pour les exons 3, 4, 5 et 12 de *MSH2* du fait des variations IVS \pm 1/2 se sont révélées inactives dans un test fonctionnel de tolérance à la méthylation. Il est intéressant de constater que, selon la région altérée, les conséquences d'un saut d'exon en phase sont très différentes. En effet, si l'exon perdu ne code pour aucun domaine fonctionnel protéique connu, il est difficile de statuer sur la fonctionnalité de la protéine résultante. Il est alors possible que la protéine conserve sa fonction, même partiellement. C'est le cas de l'exon 12 de *BRCA2* qui ne semble pas associé à un rôle indispensable pour la fonction de recombinaison homologue de la protéine *BRCA2*. Sur la base de ces résultats, la classification pathogène systématique des IVS \pm 1/2 dans cet exon a été remise en cause. En accord avec les recommandations des consortia ACMG-AMP et ENIGMA, nous proposons désormais de classer ces variations en VSI (classe 3). En revanche, si l'exon emporté code pour un domaine fonctionnel essentiel, la protéine résultante peut se voir priver d'une région indispensable à son activité, sa localisation, sa conformation ou sa stabilité, la rendant ainsi non fonctionnelle. Nos travaux ont mis en évidence que c'est le cas des exons 3, 4, 5 et 12 de *MSH2* qui sont tous associés à des domaines protéiques essentiels à l'activité de signalisation apoptotique des lésions de l'ADN de la protéine *MSH2*. Concernant l'exon 5 de *MSH2*, la variation c.942+3A>T à l'origine du saut en phase de cet exon, a déjà été établie comme variation pathogène

sur la base d'évidences de co-ségrégation avec la maladie (Mangold *et al.*, 2005). Nos données confirment donc que le saut de l'exon 5 de *MSH2*, même s'il est en phase, ne permet pas de produire une protéine fonctionnelle. Pour les exons 3, 4 et 12, nous apportons ici de nouvelles évidences quant à la nature indispensable des domaines protéiques codés par ces régions. Ainsi, d'après nos données, l'ensemble des IVS±1/2 localisées au niveau des exons 3, 4, 5 et 12 de *MSH2* demeurent des variations pathogènes.

L'exon 12 du gène *BRCA2* est donc une séquence apparemment non nécessaire d'un point de vue fonctionnel. On peut alors se demander s'il en va de même pour d'autres exons en phase du gène *BRCA2*. De façon intéressante, il a été rapporté que des IVS±1/2 localisées au niveau de l'exon 3 de *BRCA2* sont responsables du saut total en phase de cet exon dans le test fonctionnel basé sur l'utilisation de minigènes (Caputo *et al.*, 2018 ; Tubeuf *et al.*, 2020a). Ces travaux, menés en collaboration avec le réseau national GGC et le consortium international ENIGMA, ont également permis de démontrer que le saut drastique en phase de l'exon 3 de *BRCA2* ne permet pas la production d'une protéine fonctionnelle, confirmant ainsi le caractère pathogène des variations à l'origine d'un saut total de cet exon. Récemment, l'équipe du Dr Vreeswijk a mis à profit l'essai basé sur l'utilisation de cellules souches embryonnaires de souris pour confirmer l'absence de fonctionnalité de la protéine BRCA2 délétée pour l'exon 3 (Mesman *et al.*, 2020). Cette étude a étendu ses investigations aux conséquences fonctionnelles, d'une part, de certaines IVS±1/2 de *BRCA2* et, d'autre part, des délétions de chaque exon de *BRCA2*. Ces travaux ont révélé que les variations IVS±1/2 responsables du saut en phase des exons 11, 17, 19 et 26 ne permettent pas la production d'isoformes protéiques fonctionnelles. Ainsi, il semblerait que les exons en phase 11, 17, 19 et 26 codent des domaines fonctionnels essentiels à l'activité de la protéine BRCA2. Les IVS±1/2 localisées au niveau de ces exons restent donc pathogènes. A l'inverse, le maintien de la fonctionnalité de la protéine BRCA2 délétée pour l'exon 10 suggère que les IVS±1/2 à l'origine du saut en phase de cet exon pourraient voir leur classification pathogène remise en cause. Il sera important à l'avenir d'analyser l'impact de ces IVS±1/2 sur l'épissage de l'exon 10 de *BRCA2* et de déterminer leurs conséquences fonctionnelles au niveau de la protéine. De façon remarquable, les équipes du Dr Vreeswijk et du Dr Sharan ont également pu démontrer que certaines IVS±1/2 localisées cette fois dans des exons hors phase de *BRCA2* sont capables de produire des protéines en partie fonctionnelles grâce au saut en phase combiné de plusieurs exons (e.g. saut en phase des exons 4 à 7) (Biswas *et al.*, 2011 ; Thirthagiri *et al.*, 2016 ; Mesman *et al.*, 2020).

Si le paradigme des variations nulles de type IVS±1/2 à l'origine d'un saut d'exon en phase et de la production d'une protéine en partie fonctionnelle est nouveau dans les gènes de prédisposition

aux cancers, il est important de noter que ce type de mécanisme de « sauvetage » a déjà été décrit dans le contexte d'autres maladies monogéniques. Ainsi, dans le gène *DMD* qui code la protéine dystrophine, alors que des variations de type perte de fonction sont responsables d'un phénotype sévère de dystrophie musculaire de Duchenne (DMD), certaines variations $IVS\pm 1/2$ à l'origine d'un défaut d'épissage en phase ont été rapportées comme étant associées à une présentation phénotypique atténuée, décrite sous le nom de dystrophie musculaire de Becker (DMB). C'est notamment le cas de la variation *DMD* c.1603-2A>C qui est responsable du saut en phase de l'exon 14 et permet ainsi la production d'une protéine en partie fonctionnelle (Tuffery-Giraud *et al.*, 2005).

I.1.2. Les variations $IVS\pm 1/2$ à l'origine de délétions exoniques ou de rétentions introniques en phase

Les anomalies en phase induites par des variations $IVS\pm 1/2$ ne se limitent pas à des sauts d'exon. En effet, d'autres biotypes peuvent conduire à l'utilisation en phase de sites cryptiques ou alternatifs d'épissage, préexistants dans l'environnement exonique ou intronique de la mutation. Ces sites cryptiques ou alternatifs d'épissage sont susceptibles de prendre le relais à la suite de la destruction des sites constitutifs d'épissage dans le contexte muté. Les conséquences de telles anomalies d'épissage sont la production d'isoformes présentant des délétions exoniques ou des rétentions introniques en phase. Mes travaux de thèse ont apporté des preuves expérimentales de l'existence de ces événements d'épissage en phase dans le cadre de l'étude de certaines variations du gène *MSH2*. Peu de travaux dans la littérature rapportent la caractérisation de telles mutations d'épissage et, encore moins, les conséquences fonctionnelles au niveau des isoformes protéiques résultantes. Dans le cas des $IVS\pm 1/2$ à l'origine de ce type d'anomalies au niveau des exons 4, 7, 12 et 15 de *MSH2*, mes travaux de thèse ont permis de démontrer que les isoformes résultantes ne conduisent pas à la production de protéines *MSH2* fonctionnelles. Il semblerait donc que ces délétions exoniques et ces rétentions introniques en phase soient responsables de l'altération d'un domaine fonctionnel essentiel et/ou de la déstabilisation de la protéine. Il sera intéressant dans le futur de déterminer les raisons exactes de ces pertes de fonction. L'ensemble des variations $IVS\pm 1/2$ analysées ici, malgré leur effet en phase sur l'épissage, sont donc responsables de la perte de fonction de *MSH2*. Ces données confirment alors leur classification pathogène. Ces travaux suggèrent également que la protéine *MSH2*, contrairement à *BRCA2*, ne tolère pas les insertions/délétions qu'elles soient de grande ou de petite taille.

En parallèle de mes travaux de thèse, j'ai eu l'opportunité de participer à un projet collaboratif au sein du consortium international ENIGMA (pilotes : Dr de la Hoya et Dr Krieger) portant sur des variations responsables d'anomalies d'épissage en phase dans *BRCA1*, l'autre gène majeur de

prédisposition au syndrome sein-ovaire. Mes travaux ont permis de caractériser l'impact sur l'épissage de 14 variations dont 12 IVS±1/2 localisées au niveau des exons 6, 8, 13 et 14 de ce gène. Dans le test basé sur l'utilisation de minigènes, nous avons démontré que toutes ces variations IVS±1/2 favorisent l'utilisation de sites cryptiques ou alternatifs exoniques conduisant ainsi à la production d'isoformes présentant des délétions en phase de petites portions exoniques (3 à 9 nucléotides). La fonctionnalité de telles isoformes reste encore à évaluer et fait actuellement l'objet d'investigations par le consortium ENIGMA. De plus, dans ce réseau international, les données cliniques et familiales des patients porteurs sont en cours de collecte afin de statuer sur l'interprétation clinique de ces variations splicéogéniques.

I.1.3. Autres mécanismes à l'origine du maintien d'une fonction dans le contexte de variations IVS±1/2

Dans le cadre de mes travaux de thèse, nous avons abordé le cas des variations IVS±1/2 susceptibles de maintenir une fonction du fait de leur impact sur l'épissage, par le biais de modifications en phase. D'autres mécanismes en *cis* peuvent être envisagés pour expliquer un maintien de la fonction associée à certaines variations IVS±1/2.

Il est possible qu'une partie des IVS±1/2 à l'origine de défauts d'épissage hors phase permettent tout de même la production d'une protéine fonctionnelle, au moins en partie, via la production d'isoformes alternatives potentiellement fonctionnelles. Dans le gène *BRCA1*, il a notamment été montré que la variation c.[594-2A>C;641A>G] initialement classée pathogène n'était en fait pas associée à une augmentation du risque de développer un cancer, bien que cette variation soit responsable du saut hors phase de l'exon 10 (Rosenthal *et al.*, 2015 ; de la Hoya *et al.*, 2016). L'hypothèse alors proposée pour expliquer cette contradiction apparente s'appuie sur l'existence d'un épissage alternatif en phase des exons 9 et 10 ($\Delta E9-10$) qui est maintenu dans le contexte muté et qui pourrait expliquer la persistance d'une activité suffisante de *BRCA1*. Dans le futur, il sera important de vérifier la fonctionnalité de l'isoforme *BRCA1* $\Delta E9-10$ pour consolider cette explication. Un tel mécanisme de sauvetage est également envisageable dans le cas de la variation c.631+2T>G, responsable du saut hors phase de l'exon 7 de *BRCA2* et de ce fait, considérée pathogène (Pyne *et al.*, 2000). En effet, l'expression, même à de faible taux, d'une isoforme protéique *BRCA2* $\Delta E4-7$ issue d'un épissage alternatif semble permettre le maintien d'une fonction résiduelle dans le contexte muté (Biswas *et al.*, 2011 ; Thirthagiri *et al.*, 2016 ; Stauffer *et al.*, 2020 ; Mesman *et al.*, 2020). Ceci pourrait expliquer la viabilité d'individus homozygotes pour cette variation néanmoins atteints d'une anémie de Fanconi. Toutefois, il n'a pas encore été déterminé si les individus hétérozygotes pour la variation c.631+2T>G présentent un risque de développer un cancer moins élevé que celui associé à des

variations *BRCA2* clairement pathogènes. Ces données soulignent l'importance de prendre en considération la possible existence d'isoformes alternatives fonctionnelles dans l'interprétation biologique des variations. En effet, la présence de ces isoformes alternatives est susceptible de contrecarrer ou d'atténuer le caractère pathogène de certaines variations, en fonction de leur activité et de leur niveau d'expression.

Le maintien d'une fonction peut également être observé dans le contexte particulier de certaines variations IVS+2. En effet, les substitutions T>C localisées en position +2 au niveau des sites 5' d'épissage (GT>GC) ne sont pas toujours associées à un défaut d'épissage complet. Des travaux récents ont montré que, dans 15 à 18% des cas, ce type de variations permet la production plus ou moins importante (1 à 89%) de transcrits pleine longueur (Lin *et al.*, 2019 ; Chen *et al.*, 2020 ; Nix *et al.*, 2021). Ainsi, il est possible que certaines variations IVS+2 conduisent, malgré tout, à la production d'une protéine sauvage totalement fonctionnelle. Ici encore se pose la question du seuil à partir duquel la production de transcrit pleine longueur permet d'assurer un niveau de protéine fonctionnelle suffisant pour ne pas conduire à un phénotype clinique sévère. L'haploinsuffisance dépend alors du gène considéré. Dans le cas du gène *DMD*, il a été démontré que la variation c.8027+2T>C en permettant la production de seulement 10% de transcrits pleine longueur était responsable d'un phénotype atténué de dystrophie musculaire de Becker (Bartolo *et al.*, 1996).

I.2. Les variations non-sens à l'origine d'un épissage en phase et responsables du maintien d'une fonction au niveau de la protéine

I.2.1. Les variations non-sens à l'origine d'un saut d'exon en phase

Toutes les variations non-sens de l'exon 12 de *BRCA2*, répertoriées dans les bases de données mutationnelles au début de mes travaux de thèse, ont été sélectionnées afin d'étudier leur impact sur l'épissage. Parmi les 5 variations analysées, nous avons pu mettre en évidence que 4 variations sont à l'origine d'un défaut d'épissage partiel avec, d'une part, la production d'un transcrit pleine longueur porteur du PTC et, d'autre part, la production d'un transcrit avec le saut en phase de l'exon 12 ($\Delta E12$). Deux groupes de variations non-sens ont pu être distingués grâce à des analyses semi-quantitatives sur l'ARN : (i) celles avec un effet modéré sur l'épissage qui conduisent à la production minoritaire du transcrit $\Delta E12$ et (ii) celles avec un effet important sur l'épissage qui sont responsables de la production majoritaire du transcrit $\Delta E12$. Dans un contexte de variations non-sens à distance des sites d'épissage, l'impact de ces variations est très probablement le résultat direct de l'altération d'éléments régulateurs exoniques de l'épissage, *i.e.* la destruction d'un élément activateur (ESE) et/ou la création d'un élément inhibiteur (ESS). De façon intéressante, les deux variations associées aux effets les plus

drastiques sur l'épissage, les variations c.6844G>T (p.Glu2282*) et c.6901G>T (p.Glu2301*), sont localisées dans des régions riches en purines 5'-AGAA-3' qui ressemblent aux motifs décrits comme sites de liaison pour la protéine activatrice d'épissage Tra2- β 1 (Cléry *et al.*, 2011). Il est envisageable que les variations non-sens étudiées ici perturbent la liaison du facteur activateur d'épissage Tra2- β 1 en altérant ce motif et favorisent ainsi le saut de l'exon 12 de *BRCA2*. Ces anomalies d'épissage sont difficiles à appréhender par le biais des prédictions bioinformatiques basées sur les scores Δ tESRseq. En effet, dans notre jeu de données, les prédictions bioinformatiques de l'effet des variations sur l'épissage ne concordent pas toujours avec les résultats expérimentaux. Il serait donc intéressant d'envisager l'utilisation d'autres outils de prédictions *in silico* dédiés aux ESR, voir d'en combiner plusieurs, comme décrit récemment par notre équipe (Tubeuf *et al.*, 2020b), afin d'évaluer leurs capacités prédictives dans le contexte de l'exon 12 de *BRCA2*.

A l'instar des IVS \pm 1/2, les variations non-sens responsables des effets sur l'épissage les plus importants permettent la production d'une protéine *BRCA2* délétée pour l'exon 12 fonctionnelle, au moins en partie, remettant en cause leur classification pathogène. Se pose également ici la question de l'interprétation biologique et clinique des variations localisées dans cet exon particulier de *BRCA2* et présentant un effet modéré ou sans impact sur l'épissage. En effet, l'existence d'un épissage alternatif à des taux relativement important (15,9%) (Davy *et al.*, 2017) pourrait potentiellement atténuer l'effet délétère de certaines de ces variations.

Dans nos gènes d'intérêt, ce n'est pas la première fois qu'une variation non-sens est décrite comme étant à l'origine d'un saut d'exon en phase. C'est, par exemple, le cas de la variation non-sens c.5080G>T (p.Glu1694*) localisée dans l'exon 18 de *BRCA1* qui conduit au saut en phase de cet exon, apparemment de façon partielle, du fait de l'altération d'un ESE (Mazoyer *et al.*, 1998). Mais ici, le saut de cet exon emporte une séquence codant une partie du domaine BRCT, région essentielle à la fonction de *BRCA1*. Pour cette raison, cette variation conserve son caractère pathogène. Ces données soulignent l'importance de cartographier et caractériser les domaines fonctionnels des protéines afin d'interpréter finement les défauts d'épissage et d'évaluer ainsi la pathogénicité de la variation impliquée. Récemment, il a été montré que les variations non-sens *BRCA2* c.491T>A (p.Leu164*) et c.581G>A (p.Trp194*), localisées respectivement dans les exons 6 et 7 de *BRCA2*, sont à l'origine d'une augmentation de l'épissage alternatif en phase des exons 4 à 7 tandis que la variation c.9572G>A (p.Trp3191*) est responsable du saut en phase de l'exon 26 (Mesman *et al.*, 2020). Dans le test fonctionnel basé sur l'utilisation de cellules souches embryonnaires de souris, ces variations sont responsables d'une complémentation intermédiaire et d'une activité de recombinaison homologue sévèrement impactée (Mesman *et al.*, 2020). Ces données suggèrent donc que, dans le cas de ces 3

variations, la production et la fonction de ces isoformes en phase sont trop faibles pour permettre un niveau suffisant d'activité BRCA2.

De façon importante, si le paradigme des variations non-sens responsables du maintien d'une fonction par le biais de leur impact sur l'épissage est nouveau dans nos gènes d'intérêt, cela a déjà été décrit pour d'autres maladies monogéniques. Un exemple emblématique est celui de certaines variations non-sens dans le gène *DMD*. En effet, elles sont normalement responsables d'un phénotype sévère de dystrophie musculaire de Duchenne et, en fait, à l'origine d'une altération en phase de l'épissage conduisant à la production d'une protéine en partie fonctionnelle responsable d'un phénotype atténué de dystrophie musculaire de Becker (Shiga *et al.*, 1997 ; Ginjaar *et al.*, 2000 ; Tuffery-Giraud *et al.*, 2005 ; Dissert *et al.*, 2006 ; Flanigan *et al.*, 2011). C'est notamment le cas de la variation non-sens c.4250T>A (p.Leu1417*) qui est à l'origine du saut en phase de l'exon 31 de *DMD* (Tuffery-Giraud *et al.*, 2005 ; Flanigan *et al.*, 2011). Il est remarquable de constater que, dans cette maladie, un taux très faible de protéine fonctionnelle (< 5%) suffit à faire basculer le phénotype vers une forme atténuée (de Feraudy *et al.*, 2020). Des variations non-sens splicéogéniques identifiées dans d'autres gènes conduisent également à un phénotype atténué, dans le contexte notamment de la dystrophie musculaire congénitale, de l'amaurose congénitale de Leber et de l'épidermolyse bulleuse simple généralisée sévère (Di Blasi *et al.*, 2001 ; Littink *et al.*, 2010 ; Kowalewski *et al.*, 2016). Ces observations et le fait que des délétions internes en phase au niveau du gène *DMD* sont associées à des DMB ont conduit au développement de nouvelles stratégies thérapeutiques basées sur la modulation de l'épissage, notamment le saut d'exon, par l'utilisation d'oligonucléotides anti-sens (Aartsma-Rus and van Ommen, 2007 ; Pitout *et al.*, 2019). Ces molécules, constituées de courtes séquences nucléotidiques complémentaires d'une région cible du pré-ARNm, permettent de moduler spécifiquement l'épissage de façon à rétablir le cadre de lecture de transcrits *a priori* porteurs d'altérations tronquantes. Ainsi, ces agents permettent de restaurer la production d'une protéine fonctionnelle, au moins en partie. Deux molécules approuvées par la FDA en 2016 ont démontré une bonne efficacité thérapeutique dans le traitement de la DMD et de la SMA. La première molécule est l'Eteplirsén, un oligonucléotide anti-sens qui cible un élément exonique activateur de l'épissage localisé dans l'exon 51 du gène *DMD* provoquant ainsi le saut de cet exon (Lim *et al.*, 2017 ; Echevarria *et al.*, 2018). L'utilisation de cette molécule est recommandée dans le cas de patients atteints de DMD porteurs de délétions hors phase en amont de cette région du gène car, en induisant le saut de l'exon 51, elle permet de restaurer la phase de lecture et de conduire à la production de protéines dystrophines partiellement fonctionnelles. De façon remarquable, il a été constaté une perte de la marche plus lente chez des patients traités pendant 3 ans à l'Eteplirsén. La seconde molécule approuvée par la FDA est la Nusinersén qui cible un élément intronique inhibiteur de l'épissage localisé

dans l'intron 7 du gène *SMN2* et augmente ainsi l'inclusion de l'exon 7, permettant la production d'une protéine SMN pleine longueur fonctionnelle (Paton, 2017). L'utilisation de cette molécule est recommandée dans le cas de patients atteints de SMA, maladie provoquée par une perte bi-allélique du gène *SMN1* afin de compléter la déficience en protéines SMN. Les patients traités ont démontré de nombreuses améliorations dans la fonction motrice. Au vu de l'existence, dans certains gènes de prédisposition aux cancers (*BRCA1* et *BRCA2*), des possibilités de sauvetage par le biais de l'épissage, on peut se demander dans quelle mesure de telles stratégies thérapeutiques pourraient être adaptées dans le futur afin de prévenir le développement de cancers chez les individus porteurs de variations perte de fonction.

Si les variations non-sens sont capables d'induire des modifications de l'épissage par le biais de saut en phase, elles peuvent également conduire à des délétions exoniques en phase. En parallèle de mes travaux de thèse portant sur *BRCA2* et *MSH2*, j'ai eu l'opportunité d'étudier une variation non-sens dans le gène *APC* impliqué dans la polypose adénomateuse familiale (PAF). Il s'agit de la variation c.288T>A (p.Tyr96*) localisée dans l'exon 3. De façon intéressante, cette variation a été identifiée chez 33 individus issus de 7 familles différentes dont 28 ont été diagnostiqués pour un cancer colorectal à un âge moyen de 61,7 ans (Ibrahim *et al.*, 2014). Or, dans les formes classiques de polypose adénomateuse familiale, l'âge moyen de cancer colorectal est généralement autour de 40 ans. Cette variation est décrite comme associée à un phénotype atténué. Nous avons alors émis l'hypothèse selon laquelle cette observation pourrait être la conséquence de l'impact de la variation sur l'épissage. Nous avons pu mettre en évidence, à l'aide du test minigène et à partir de l'ARN de patients, que cette variation engendre majoritairement la production d'un transcrit pleine longueur porteur d'un PTC mais elle est également responsable de l'expression minoritaire d'un transcrit délété pour les 69 premiers nucléotides de l'exon 3. Cette anomalie en phase est la conséquence de la création d'un site 3' d'épissage exonique dans le contexte muté. De futures investigations seront nécessaires pour déterminer si l'isoforme protéique correspondante (délétion interne de 23 acides aminés) est fonctionnelle, ce qui pourrait expliquer pourquoi cette variation non-sens est associée à une forme atténuée de PAF.

Plus largement, le mécanisme de sauvetage des mutations non-sens par le processus d'épissage pourrait également exister en dehors d'un contexte de maladie monogénique. En effet, il a été estimé que, dans la population générale, chaque individu est porteur en moyenne de 100 variations présumées pertes de fonction, dont 20% à l'état homozygote (MacArthur *et al.*, 2012). Certaines de ces variations pourraient également permettre le maintien d'une fonction par le biais de leur impact direct sur l'épissage ou du fait d'un épissage alternatif (Jagannathan and Bradley, 2016).

L'existence potentielle d'un mécanisme de sauvetage associé à certaines variations non-sens splicéogéniques peut également être à l'origine de résultats inattendus dans le cadre de certains travaux de recherche. C'est notamment le cas d'approches de génétique inverse qui mettent à profit des modèles animaux pour établir un lien de causalité entre la perte de fonction d'un gène et l'expression d'un phénotype spécifique. Il a ainsi été rapporté que l'observation inattendue d'un phénotype normal de certains poissons zèbres mutants supposés *knock out* était le fait des propriétés splicéogéniques de variations non-sens introduites dans le génome de ces animaux (Anderson *et al.*, 2017). Il est désormais conseillé de prendre en considération ce possible écueil dans la sélection des variations *bona fide* perte de fonction pour l'établissement de modèles animaux.

1.2.2. Autres mécanismes à l'origine d'un maintien d'une fonction dans le contexte de variations non-sens

Dans le cadre de mes travaux de thèse, nous avons abordé le cas des variations non-sens susceptibles de maintenir une fonction du fait de leur impact direct sur l'épissage. D'autres mécanismes en *cis* peuvent contourner la perte totale de fonction dans le contexte de ce type de variations, en particulier du fait (i) de l'épissage alternatif physiologique de l'exon porteur du PTC, (ii) de la localisation C-terminale du PTC, (iii) de la localisation N-terminale du PTC ou (iv) de la translecture.

Tout comme pour les IVS \pm 1/2, il est envisageable qu'une partie des variations non-sens permettent tout de même la production d'une protéine fonctionnelle, au moins en partie, du fait d'un épissage alternatif en phase de l'exon porteur du PTC. Cet évènement permet l'élimination du codon stop prématuré et potentiellement la production d'une isoforme protéique fonctionnelle. Cela pourrait être le cas des variations non-sens dans l'exon 12 de *BRCA2* sans impact sur l'épissage puisque cet exon non essentiel est soumis à un épissage alternatif (Li *et al.*, 2009 ; Fackenthal *et al.*, 2016 ; Davy *et al.*, 2017). De futures investigations seront nécessaires pour déterminer si le niveau d'expression de cette isoforme alternative permet une activité *BRCA2* suffisante. On peut se poser la même question pour des variations non-sens splicéogéniques localisées dans d'autres exons de *BRCA2* (*i.e.* Δ E4-7) et de *BRCA1* (*i.e.* Δ E9-10), puisque ces sauts combinés d'exons sont également alternatifs et susceptibles de conduire à la production d'une isoforme fonctionnelle (Thirthagiri *et al.*, 2016 ; Stauffer *et al.*, 2020 ; de la Hoya *et al.*, 2016 ; Mesman *et al.*, 2020).

D'autres évènements permettent de contourner la perte de fonction engendrée par les variations non-sens. Tout d'abord, les PTC localisés dans la partie C-terminale peuvent entraîner la perte d'une séquence non essentielle à la fonction protéique. On peut notamment citer le cas de la

variation non-sens *BRCA2* c.9976A>T (p.Lys3326*) initialement supposée perte de fonction. Cette variation a fait l'objet d'une reclassification en variation non pathogène sur la base notamment de sa fréquence élevée dans la population générale (Mazoyer *et al.*, 1996). Localisée à l'extrémité C-terminale de la protéine, cette variation entraîne la production d'une protéine tronquée pour les 93 derniers acides aminés. Malgré cette délétion, il a été montré que la protéine tronquée correspondante est fonctionnelle (Wu *et al.*, 2005). Toutefois, et de façon intrigante, cette variation est associée à un risque très faible de cancers du sein et de l'ovaire (Odd ratio = 1,28 et 1,26) (Meeks *et al.*, 2015). Il est donc possible qu'il s'agisse d'une variation hypomorphe, mais cela reste encore à clarifier.

Les variations non-sens à l'origine d'un PTC dans la partie N-terminale peuvent également permettre le maintien d'une fonction dans certains cas. Cela est possible du fait de la ré-initiation de la traduction en aval du codon stop prématuré du fait de la présence d'un codon d'initiation alternatif en phase. En conséquence, seulement quelques acides aminés seront délétés au début de la protéine. Ce mécanisme est bien connu et décrit de nombreuses fois dans la littérature. On peut notamment citer le cas de la variation non-sens c.109C>T (p.Arg37*) dans le gène *ATRX* qui aurait dû engendrer un phénotype léthal. L'utilisation d'un codon d'initiation alternatif 3 acides aminés en aval du PTC permet de produire environ 20% de protéine *ATRX* fonctionnelle, délétée pour les 40 premiers acides aminés, avec pour conséquence un syndrome ATR-X moins sévère (Howard *et al.*, 2004). Dans le gène *APC*, la ré-initiation de la traduction au niveau d'un codon AUG en position 184 serait également responsable de formes atténuées de PAF observées dans le contexte de variations non-sens localisées à l'extrémité N-terminale de la protéine (Heppner Goss *et al.*, 2002). A notre connaissance, dans les gènes *BRCA* et *MMR*, il n'a pas été décrit de cas de variations non-sens localisées dans la région N-terminale qui seraient associées à un phénotype atténué. Cette possibilité ne peut toutefois être exclue, notamment pour le gène *MSH2*. En effet, il a été rapporté l'existence d'une variation localisée sur la première base du codon d'initiation de la traduction du gène *MSH2* (c.1A>C) qui serait associée à une pénétrance bien plus faible que celles établies pour les variations pathogènes classiquement associées à un syndrome de Lynch (Cyr *et al.*, 2012 ; Rosenthal *et al.*, 2015). La présence d'un codon d'initiation de la traduction alternatif 26 acides aminés en aval du codon AUG constitutif permettrait la production d'une protéine fonctionnelle. Sur la base de ces données, on peut alors se questionner sur le réel impact des variations non-sens localisées en amont de ce codon d'initiation alternatif.

Un autre phénomène est susceptible d'expliquer pourquoi certaines variations non-sens ne provoquent pas une perte totale de fonction de la protéine. Il met en jeu le processus de translecture du codon stop. Ce mécanisme permet l'incorporation d'un acide aminé à la place du codon stop

prématuré, au cours de la traduction (Dabrowski *et al.*, 2015). Le phénomène de translecture spontanée a été largement décrit chez les virus, les bactéries et les levures (Mottagui-Tabar and Isaksson, 1998 ; Rospert *et al.*, 2005 ; Schueren and Thoms, 2016). Il dépend directement du contexte nucléotidique et de la nature du codon stop. A ce jour, le phénomène de translecture a très peu été décrit dans un contexte de maladies mendéliennes. On peut toutefois citer le cas de la variation *LAMA3* c.2827C>T (p.Arg943*), responsable d'un phénotype atténué d'épidermolyse bulleuse jonctionnelle. Dans le contexte muté, le transcrit porteur du PTC n'est pas dégradé par le NMD et conduirait à la production d'une protéine pleine longueur, grâce au mécanisme de translecture, permettant ainsi l'expression d'une forme atténuée de la maladie (Pacho *et al.*, 2011). Un autre cas de translecture en lien avec une maladie mendélienne a été décrit dans le contexte de plusieurs variations non-sens du gène *F9* impliqué dans l'hémophilie B, caractérisée par un déficit du facteur IX. En présence de ces variations, une faible quantité de protéines pleine longueur a pu être observée dans le sang des patients du fait d'un phénomène de translecture spontané (Pinotti *et al.*, 2012). Cependant, à l'inverse du premier exemple, la quantité de protéines pleine longueur ainsi produites n'est pas suffisante pour modifier le phénotype de la maladie. Il est intéressant de noter que, depuis une vingtaine d'années, l'utilisation de molécules spécifiques, telles que les antibiotiques de la classe des aminoglycosides, favorisant la translecture des codons stop prématurés, a été considérée comme une stratégie thérapeutique prometteuse (Keeling *et al.*, 2014). Un autre agent pharmacologique, l'ataluren, qui stimule également la translecture des 3 codons stop, a montré une efficacité importante pour certains patients atteints de dystrophie musculaire de Duchenne et porteurs de PTC dans le gène *DMD* (Peltz *et al.*, 2013 ; Mercuri *et al.*, 2020). Ce traitement a en effet retardé la progression de la maladie. Il faut toutefois souligner que le phénomène de translecture est en général peu efficace, qu'il soit spontané ou induit par des drogues (Schueren and Thoms, 2016 ; Rodnina *et al.*, 2020). Cette stratégie thérapeutique ne peut donc être envisagée aujourd'hui que dans le contexte de pathologies pour lesquelles une amélioration du phénotype est possible du fait de l'expression, même faible, d'une protéine fonctionnelle.

II. La problématique de l'interprétation clinique des variations hypomorphes

Les variations hypomorphes responsables de la production d'une protéine en partie fonctionnelle, telles que certaines variations IVS±1/2 et non-sens de l'exon 12 de *BRCA2*, représentent un nouveau défi particulièrement important en termes d'interprétation clinique dans un contexte d'oncogénétique médicale. Dans les gènes *BRCA*, peu de variations ont été précédemment décrites comme responsables d'un niveau intermédiaire d'activité. Certaines d'entre elles se sont révélées être à l'origine d'une atténuation de l'expression clinique du syndrome sein-ovaire avec une diminution du

risque associé de développer un cancer. L'exemple le plus emblématique est celui de la mutation faux-sens *BRCA1* c.5096G>A (p.Arg1699Gln) qui a fait l'objet de nombreuses investigations (Lovelock *et al.*, 2007 ; Spurdle *et al.*, 2012 ; Shimelis *et al.*, 2017 ; Moghadasi *et al.*, 2018). Il a ainsi été montré que cette variation était responsable d'une activité intermédiaire dans différents essais fonctionnels protéiques (Lovelock *et al.*, 2007 ; Shimelis *et al.*, 2017 ; Spurdle *et al.*, 2018). La protéine mutée présente un domaine BRCT altéré, à l'origine de la perte partielle de fonction (Lovelock *et al.*, 2007 ; Spurdle *et al.*, 2018). Afin d'appréhender le risque de cancer lié à cette variation hypomorphe, des études de co-ségrégation ont été entreprises. Une première étude basée sur l'analyse de 30 familles a révélé que cette variation était responsable d'un risque cumulé de cancers du sein et de l'ovaire estimé à 24% à 70 ans, risque bien inférieur à celui estimé à 68% pour les variations associées à une perte totale de fonction de *BRCA1* (Spurdle *et al.*, 2012). Plus récemment, une étude à plus grande échelle conduite au sein du consortium ENIGMA et incluant 129 familles a permis de déterminer de façon robuste les risques de cancers de sein et de l'ovaire à 70 ans pour cette variation, estimés à 20% et 6%, respectivement (Moghadasi *et al.*, 2018).

Qu'en est-il alors de la classification de cette variation ? Les faibles risques associés à cette variation hypomorphe de *BRCA1* ne permettent théoriquement pas de la classer dans le système aujourd'hui établi pour rendre compte de la pathogénicité des variations dans les gènes *BRCA*. En effet, ce système est binaire. Il permet la distinction entre, d'une part, les variations neutres sans impact biologique et sans risque (classe 1) et, d'autre part, les variations pathogènes avec perte totale de fonction et associées à un risque élevé de développer la maladie (classe 5). La mise en évidence de variations hypomorphes responsables d'un risque plus faible suggère qu'il existe en fait un continuum de la sévérité de l'impact biologique des variations, avec pour conséquence des risques plus ou moins importants de développer la maladie. La pénétrance pourrait donc être variable en fonction des différentes variations identifiées au sein d'un même gène, avec des implications importantes pour la prise en charge médicale des patients porteurs et de leurs apparentés.

Aujourd'hui des réflexions sont menées pour adapter la prise en charge médicale des individus porteurs de variations hypomorphes à risque modéré dans les gènes *BRCA*. Dans le cas de la variation *BRCA1* c.5096G>A (p.Arg1699Gln) le consortium ENIGMA a recommandé une surveillance mammaire pour les femmes porteuses sur la base d'une mammographie annuelle à partir de 40 ans et une hystérectomie doit être considérée en fonction de l'histoire familiale (Moghadasi *et al.*, 2018). Ces recommandations spécifiques s'appuient sur celles précédemment émises dans un contexte de syndrome sein-ovaire pour les gènes dont la perte de fonction est responsable d'un risque modéré de cancer, tel que *CHEK2* (Easton *et al.*, 2015 ; Tung *et al.*, 2016).

L'existence d'autres variations faux-sens hypomorphes de *BRCA1* et *BRCA2*, associées à un risque modéré, a été suggérée par plusieurs études (Domchek *et al.*, 2013 ; Spurdle *et al.*, 2012 ; Shimelis *et al.*, 2017 ; Monteiro *et al.*, 2020). Cependant, ces investigations se confrontent souvent à deux problèmes majeurs : (i) la non-concordance éventuelle des données générées dans différents essais fonctionnels pour la même variation, et (ii) la difficulté d'estimer le risque du fait de la rareté de la variation. Il est aujourd'hui impossible d'établir une corrélation directe, pour une variation donnée, entre le niveau de fonctionnalité mesuré dans un essai spécifique et le risque associé de cancers. De futures investigations seront nécessaires pour évaluer la possibilité d'une calibration des tests fonctionnels dans ce sens. Elles devraient également permettre d'appréhender les seuils d'haploinsuffisance minimaux associés à des variations *bona fide* pathogènes qui confèrent des risques élevés ainsi que celles conférant des risques intermédiaires.

Toutes ces questions se posent également dans le contexte de variations splicéogéniques à effet partiel, c'est à dire qui conduisent à la production de transcrits aberrants hors phase mais également à celle des transcrits pleine longueur, à des niveaux plus ou moins importants. C'est notamment le cas de la variation c.5407-25T>A localisée dans l'intron 22 de *BRCA1* qui est responsable de la production de deux transcrits : l'un, présentant le saut hors phase de l'exon 23 et l'autre, étant pleine longueur (Hoberg-Vetti *et al.*, 2020). De façon intéressante, cette anomalie a été identifiée dans 20 familles associées à un risque intermédiaire de cancer, probablement du fait de la production résiduelle de l'isoforme pleine longueur. Dans les gènes *BRCA* et *MMR*, de nombreuses variations aujourd'hui classées VSI ont été identifiées comme étant à l'origine d'un effet partiel sur l'épissage (Tournier *et al.*, 2008 ; Théry *et al.*, 2011 ; Gaildrat *et al.*, 2012 ; Houdayer *et al.*, 2012 ; Di Giacomo *et al.*, 2013 ; Soukariéh *et al.*, 2016 ; Tubeuf *et al.*, 2020a). La pathogénicité de telles variations splicéogéniques potentiellement hypomorphes dépend des niveaux d'expression des transcrits pleine longueur sauvages. Se pose alors la question du seuil. Très récemment, une étude de notre groupe à laquelle j'ai contribué a permis de montrer, en utilisant comme modèle l'exon 3 de *BRCA2*, que des variations splicéogéniques à effet partiel n'étaient pas pathogènes si elles permettaient la production d'au moins 60% de transcrits pleine longueur sauvages (Tubeuf *et al.*, 2020a). Ces travaux représentent la première calibration de la pathogénicité des variations splicéogéniques à effet partiel dans ce gène. Dans le futur, ce type d'approches devrait contribuer à la classification des très nombreuses VSI splicéogéniques à effet partiel identifiées dans les gènes *BRCA* et *MMR*. Là encore, il sera indispensable d'intégrer des méthodes quantitatives précises pour l'évaluation des taux de transcrits.

Pour ce qui est des variations splicéogéniques $IVS\pm 1/2$ et non-sens de l'exon 12 de *BRCA2* étudiées dans le cadre de mes travaux de thèse, la difficulté d'interprétation clinique est encore plus complexe. En effet, d'une part, ces variations produisent des taux variables de transcrit $\Delta E12$ et donc de protéine $\Delta E12$, du fait de leur impact différentiel sur l'épissage et, d'autre part, cette protéine présente une activité réduite par rapport à la protéine *BRCA2* sauvage. Le test fonctionnel basé sur l'utilisation des cellules souches embryonnaires de souris nous a permis d'appréhender l'impact combiné de ces variations sur l'épissage et la fonction de la protéine dans le contexte de ce type de variations. Les résultats obtenus indiquent que certaines variations présentent une activité similaire à celle mesurée pour des variations neutres tandis que d'autres montrent une activité plus faible, toutefois supérieure à celle mesurée pour des variations pathogènes. L'analyse des données cliniques et familiales collectées dans le réseau national des laboratoires d'oncogénétique, pour les patients porteurs de ces variations splicéogéniques hypomorphes de l'exon 12 de *BRCA2*, n'a pas permis d'établir les risques de cancers associés à ce type de variations. De futures investigations, basées sur des études de co-ségrégation et des analyses de cas-contrôles à l'échelle internationale seront nécessaires pour estimer leurs pénétrances respectives. La difficulté dans ce type d'approches est toutefois la rareté de chaque variation. Afin de contourner le problème, il pourra être envisagé d'agréger les informations obtenues pour des variations présentant le même impact biologique, comme cela a été fait dans le cadre du projet sur l'exon 3 de *BRCA2* (Caputo *et al.*, 2018). Ceci implique l'intégration de données qualitatives et quantitatives issues de plusieurs tests fonctionnels calibrés.

A l'ère de la médecine génomique de précision, l'interprétation biologique et clinique des variations génétiques identifiées dans le génome des patients a des implications majeures en termes de diagnostic, de pronostic, de prévention mais également dans les choix thérapeutiques. Dans le domaine de l'oncogénétique, mes travaux de thèse mettent en exergue la complexité de cette interprétation, avec la nécessité de distinguer les variations responsables d'une perte totale de fonction des gènes de prédisposition aux cancers de celles qui sont hypomorphes, responsables potentiellement d'un risque intermédiaire. Ces considérations impliquent d'appréhender, de façon qualitative et quantitative, l'impact biologique des variations au niveau de l'ARN et de la protéine, à l'aide d'approches complémentaires ayant fait l'objet de calibration. Dans ce contexte, le développement récent de méthodes à très haut débit semble très prometteur et présage l'accumulation de données fonctionnelles en masse pour toute substitution (exoniques et introniques flanquantes) dans des gènes impliqués dans des maladies, avant même l'identification de ces variations chez les patients. Ce type de démarches a déjà commencé, en tout cas en partie, dans certains gènes dont *BRCA1* (Findlay *et al.*, 2018) et *MSH2* (Jia *et al.*, 2021), ce qui laisse espérer des avancées majeures en génétique médicale dans un futur proche.

Références bibliographiques

- Aartsma-Rus A, van Ommen G-JB. Antisense-mediated exon skipping: A versatile tool with therapeutic and research applications. *RNA*. 2007;13:1609–24.
- Abramowicz and Gos. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59:253–68.
- Acedo A, Hernández-Moro C, Curiel-García Á, Díez-Gómez B, Velasco EA. Functional Classification of *BRCA2* DNA Variants by Splicing Assays in a Large Minigene with 9 Exons. *Human Mutation*. 2015;36:210–21.
- Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, et al. Comprehensive splicing functional analysis of DNA variants of the *BRCA2* gene by hybrid minigenes. *Breast Cancer Res*. 2012;14:R87.
- Adamson AW, Beardsley DI, Kim W-J, Gao Y, Baskaran R, Brown KD. Methylator-induced, Mismatch Repair-dependent G₂ Arrest Is Activated through Chk1 and Chk2. *MBoC*. 2005;16:1513–26.
- Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. 2015;52.
- Agata S, De Nicolo A, Chieco-Bianchi L, D'Andrea E, Menin C, Montagna M. The *BRCA2* sequence variant IVS19+1G-->A leads to an aberrant transcript lacking exon 19. *Cancer Genet Cytogenet*. 2003;141:175–6.
- Ahmadian A, Ehn M, Hober S. Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta*. 2006;363:83–94.
- Anczuków O, Buisson M, Léoné M, Coutanson C, Lasset C, Calender A, et al. *BRCA2* Deep Intronic Mutation Causing Activation of a Cryptic Exon: Opening toward a New Preventive Therapeutic Strategy. *Clin Cancer Res*. 2012;18:4903–9.
- Anderson JL, Mulligan TS, Shen M-C, Wang H, Scahill CM, Tan FJ, et al. mRNA processing in mutant zebrafish lines generated by chemical and CRISPR-mediated mutagenesis produces unexpected transcripts that escape nonsense-mediated decay. Mullins MC, editor. *PLoS Genet*. 2017;13:e1007105.
- André T, Shiu K-K, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N Engl J Med*. 2020;383:2207–18.
- Angeli D, Salvi S, Tedaldi G. Genetic Predisposition to Breast and Ovarian Cancers: How Many and Which Genes to Test? *IJMS*. 2020;21:1128.
- Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, et al. Breast-Cancer Risk in Families with Mutations in *PALB2*. *N Engl J Med*. 2014;371:497–506.
- Aravind L. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Research*. 1999;27:1223–42.
- Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, et al. Classifying *MLH1* and *MSH2* variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat*. 2009;30:757–70.
- Arun B, Akar U, Gutierrez-Barrera AM, Hortobagyi GN, Ozpolat B. The PARP inhibitor AZD2281 (Olaparib) induces autophagy/mitophagy in *BRCA1* and *BRCA2* mutant breast cancer cells. *International Journal of Oncology*. 2015;47:262–8.
- Badie S, Escandell JM, Bouwman P, Carlos AR, Thanasoula M, Gallardo MM, et al. *BRCA2* acts as a *RAD51* loader to facilitate telomere replication and capping. *Nat Struct Mol Biol*. 2010;17:1461–9.
- Baert-Desurmont S, Coutant S, Charbonnier F, Macquere P, Lecoquierre F, Schwartz M, et al. Optimization of the diagnosis of inherited colorectal cancer using NGS and capture of exonic and intronic sequences of panel genes. *Eur J Hum Genet*. 2018;26:1597–602.
- Baralle D. Splicing in action: assessing disease causing sequence changes. *Journal of Medical Genetics*. 2005;42:737–48.
- Baralle D, Buratti E. RNA splicing in human disease and in the clinic. *Clinical Science*. 2017;131:355–68.
- Baralle D, Lucassen A, Buratti E. Missed threads: The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep*. 2009;10:810–6.
- Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol*. 2017;18:437–51.

- Baralle M, Baralle D, De Conti L, Mattocks C, Whittaker J, Knezevich A, et al. Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. *J Med Genet*. 2003;40:220–2.
- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. *Nature*. 2010;465:53–9.
- Barmada SJ. Linking RNA Dysfunction and Neurodegeneration in Amyotrophic Lateral Sclerosis. *Neurotherapeutics*. 2015;12:340–51.
- Bartolo C, Papp AC, Snyder PJ, Sedra MS, Burghes AH, Hall CD, et al. A novel splice site mutation in a Becker muscular dystrophy patient. *Journal of Medical Genetics*. 1996;33:324–7.
- Bębenek A, Ziuzia-Graczyk I. Fidelity of DNA replication—a matter of proofreading. *Curr Genet*. 2018;64:985–96.
- Berget SM. Exon Recognition in Vertebrate Splicing. *Journal of Biological Chemistry*. 1995;270:2411–4.
- Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA*. 1977;74:3171–5.
- Berk AJ. Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci USA*. 2016;113:801–5.
- Bernier FP, Caluseriu O, Ng S, Schwartzenruber J, Buckingham KJ, Innes AM, et al. Haploinsufficiency of SF3B4, a Component of the Pre-mRNA Spliceosomal Complex, Causes Nager Syndrome. *The American Journal of Human Genetics*. 2012;90:925–33.
- Bérout C, Collod-Bérout G, Boileau C, Soussi T, Junien C. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. *Hum Mutat*. 2000;15:86–94.
- Bérout C, Letovsky SI, Braastad CD, Caputo SM, Beaudoux O, Bignon YJ, et al. BRCA Share: A Collection of Clinical BRCA Gene Variants: HUMAN MUTATION. *Human Mutation*. 2016;37:1318–28.
- Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*. 2014;511:362–5.
- Bienstock RJ, Darden T, Wiseman R, Pedersen L, Barrett JC. Molecular modeling of the amino-terminal zinc ring domain of BRCA1. *Cancer Res*. 1996;56:2539–45.
- Biswas K, Das R, Alter BP, Kuznetsov SG, Stauffer S, North SL, et al. A comprehensive functional characterization of BRCA2 variants associated with Fanconi anemia using mouse ES cell-based assay. *Blood*. 2011;118:2430–42.
- Biswas K, Das R, Eggington JM, Qiao H, North SL, Stauffer S, et al. Functional evaluation of BRCA2 variants mapping to the PALB2-binding and C-terminal DNA-binding domains using a mouse ES cell-based assay. *Human Molecular Genetics*. 2012;21:3993–4006.
- Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YYC, Codlin S, et al. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Research*. 2014;24:1169–79.
- Boland CR. Familial Colonic Cancer Without Antecedent Polyposis. *Ann Intern Med*. 1984;100:700.
- Brandão RD, Mensaert K, López-Perolio I, Tserpelis D, Xenakis M, Lattimore V, et al. Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *Int J Cancer*. 2019;145:401–14.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68:394–424.
- Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proceedings of the National Academy of Sciences*. 1978;75:4853–7.
- Brewington JJ, Filbrandt ET, LaRosa FJ, Moncivaiz JD, Ostmann AJ, Strecker LM, et al. Brushed nasal epithelial cells are a surrogate for bronchial epithelial CFTR studies. *JCI Insight*. 2018;3:e99385.
- Briese M, Haberman N, Sibley CR, Faraway R, Elser AS, Chakrabarti AM, et al. A systems view of spliceosomal assembly and branchpoints with iCLIP. *Nat Struct Mol Biol*. 2019;26:930–40.
- Broca P (1824-1880). *Traité des tumeurs, par Paul Broca...* :624.
- Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol*. 2009;16:107–13.

- Brown KD, Rathi A, Kamath R, Beardsley DI, Zhan Q, Mannino JL, et al. The mismatch repair system is required for S-phase checkpoint activation. *Nat Genet.* 2003;33:80–4.
- Brzovic PS, Rajagopal P, Hoyt DW, King M-C, Klevit RE. Structure of a BRCA1– BARD1 heterodimeric RING–RING complex. *nature structural biology.* 2001;8:5.
- Bubien V, Bonnet F, Brouste V, Hoppe S, Barouk-Simonet E, David A, et al. High cumulative risks of cancer in patients with *PTEN* hamartoma tumour syndrome. *J Med Genet.* 2013;50:255–63.
- Buckanovich RJ, Posner JB, Darnell RB. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron.* 1993;11:657–72.
- Buecher B, Le Mentec M, Doz F, Bourdeaut F, Gauthier-Villars M, Stoppa-Lyonnet D, et al. Syndrome CMMRD (déficience constitutionnelle des gènes MMR) : bases génétiques et aspects cliniques. *Bulletin du Cancer.* 2019;106:162–72.
- Bujalkova M, Zavodna K, Krivulcik T, Ilencikova D, Wolf B, Kovac M, et al. Multiplex SNaPshot Genotyping for Detecting Loss of Heterozygosity in the Mismatch-Repair Genes MLH1 and MSH2 in Microsatellite-Unstable Tumors. *Clinical Chemistry.* 2008;54:1844–54.
- Burset M. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Research.* 2000;28:4364–75.
- Busslinger M, Moschonas N, Flavell RA. Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell.* 1981;27:289–98.
- Caminsky NG, Mucaki EJ, Rogan PK. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Res.* 2014;3:282.
- Canson D, Glubb D, Spurdle AB. Variant effect on splicing regulatory elements, branchpoint usage, and pseudoexonization: Strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Human Mutation.* 2020;41:1705–21.
- Caputo S, Benboudjema L, Sinilnikova O, Rouleau E, Bérout C, Lidereau R, et al. Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. *Nucleic Acids Research.* 2012;40:D992–1002.
- Caputo SM, Léone M, Damiola F, Ehlen A, Carreira A, Gaidrat P, et al. Full in-frame exon 3 skipping of BRCA2 confers high risk of breast and/or ovarian cancer. *Oncotarget.* 2018;9:17334–48.
- Carr K. Nobel goes to discoverers of “split genes.” *Nature.* 1993;365:597.
- Carreira A, Hilario J, Amitani I, Baskin RJ, Shivji MKK, Venkitaraman AR, et al. The BRC Repeats of BRCA2 Modulate the DNA-Binding Selectivity of RAD51. *Cell.* 2009;136:1032–43.
- Cartegni L. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Research.* 2003;31:3568–71.
- Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002;3:285–98.
- Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR. Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, SMN1 and SMN2. *The American Journal of Human Genetics.* 2006;78:63–77.
- Cartegni L, Krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet.* 2002;30:377–84.
- Castéra L, Harter V, Muller E, Krieger S, Goardon N, Ricou A, et al. Landscape of pathogenic variations in a panel of 34 genes and cancer risk estimation from 5131 HBOC families. *Genet Med.* 2018;20:1677–86.
- Castéra L, Krieger S, Rousselin A, Legros A, Baumann J-J, Bruet O, et al. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur J Hum Genet.* 2014;22:1305–13.
- Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, et al. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet.* 2008;40:1416–25.
- Caux-Moncoutier V, Pagès-Berhouet S, Michaux D, Asselain B, Castéra L, De Pauw A, et al. Impact of BRCA1 and BRCA2 variants on splicing: clues from an allelic imbalance study. *Eur J Hum Genet.* 2009;17:1471–80.

- Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology*. 2016;26:52–64.
- Cech TR. The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell*. 1986;44:207–10.
- Chakraborty U, Alani E. Understanding how mismatch repair proteins participate in the repair/anti-recombination decision. Mortensen U, editor. *FEMS Yeast Research*. 2016;16:fow071.
- Chang Y-F, Imam JS, Wilkinson MF. The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu Rev Biochem*. 2007;76:51–74.
- Chao EC. Molecular models for the tissue specificity of DNA mismatch repair-deficient carcinogenesis. *Nucleic Acids Research*. 2006;34:840–52.
- Chappuis PO, Nethercot V, Foulkes WD. Clinico-pathological characteristics of BRCA1- and BRCA2-related breast cancer. *Semin Surg Oncol*. 2000;18:287–95.
- Charbonnier F, Martin C, Scotte M, Sibert L, Moreau V, Frebourg T. Alternative splicing of MLH1 messenger RNA in human normal cells. *Cancer Res*. 1995;55:1839–41.
- Chen C-C, Feng W, Lim PX, Kass EM, Jasin M. Homology-Directed Repair and the Role of BRCA1, BRCA2, and Related Proteins in Genome Integrity and Cancer. *Annu Rev Cancer Biol*. 2018;2:313–36.
- Chen J-M, Lin J-H, Masson E, Liao Z, Férec C, Cooper DN, et al. The Experimentally Obtained Functional Impact Assessments of 5' Splice Site GT>GC Variants Differ Markedly from Those Predicted. *CG*. 2020;21:56–66.
- Chen X, Truong T-TN, Weaver J, Bove BA, Cattie K, Armstrong BA, et al. Intronic alterations in *BRCA1* and *BRCA2* : effect on mRNA splicing fidelity and expression. *Hum Mutat*. 2006;27:427–35.
- Chiang H-L, Wu J-Y, Chen Y-T. Identification of functional single nucleotide polymorphisms in the branchpoint site. *Hum Genomics*. 2017;11:27.
- Chillon M, Dork T, Casals T, Gimenez J, Fonknechten N, Will K, et al. A Novel Donor Splice Site in Intron I I of the CFTR Gene, Created by Mutation 1811 + I .6kbA-G, Produces a New Exon: High Frequency in Spanish Cystic Fibrosis Chromosomes and Association with Severe Phenotype. *Am J Hum Genet*. 1995;7.
- Chow LT, Gelinis RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*. 1977;12:1–8.
- Clarke CL, Sandle J, Jones AA, Sofronis A, Patani NR, Lakhani SR. Mapping loss of heterozygosity in normal human breast cells from BRCA1/2 carriers. *Br J Cancer*. 2006;95:515–9.
- Cléry A, Jayne S, Benderska N, Dominguez C, Stamm S, Allain FH-T. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2- β 1. *Nat Struct Mol Biol*. 2011;18:443–50.
- Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *Eng C*, editor. *PLoS Genet*. 2018;14:e1007752.
- Coady TH, Lorson CL. SMN in spinal muscular atrophy and snRNP biogenesis: SMN in spinal muscular atrophy and snRNP biogenesis. *WIREs RNA*. 2011;2:546–64.
- Cobb M. Who discovered messenger RNA? *Current Biology*. 2015;25:R526–32.
- Coelho MB, Smith CWJ. Regulation of alternative pre-mRNA splicing. *Methods Mol Biol*. 2014;1126:55–82.
- Coffin E, Dhooge M, Abou Ali E, Dermine S, Lavole J, Palmieri LJ, et al. Syndrome de LYNCH : identification et prise en charge. *La Presse Médicale*. 2019;48:904–14.
- Cohen SN, Chang ACY, Boyer HW, Helling RB. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences*. 1973;70:3240–4.
- Cohen-Haguenaer O. Prédisposition héréditaire au cancer du sein (1): Génétique. *Med Sci (Paris)*. 2019;35:138–51.
- Coleman RL, Oza AM, Lorusso D, Aghajanian C, Oaknin A, Dean A, et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390:1949–61.
- Colombo M, Blok MJ, Whiley P, Santamaría M, Gutiérrez-Enríquez S, Romero A, et al. Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the

- ENIGMA consortium. *Human Molecular Genetics*. 2014;23:3666–80.
- Colombo M, Lòpez-Perolio I, Meeks HD, Caleca L, Parsons MT, Li H, et al. The *BRCA2* c.68-7T > A variant is not pathogenic: A model for clinical calibration of spliceogenicity. *Human Mutation*. 2018;39:729–41.
- Colussi C, Parlanti E, Degan P, Aquilina G, Barnes D, Macpherson P, et al. The Mammalian Mismatch Repair Pathway Removes DNA 8-oxodGMP Incorporated from the Oxidized dNTP Pool. *Current Biology*. 2002;12:912–8.
- Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods*. 2005;37:331–40.
- Corvelo A, Hallegger M, Smith CWJ, Eyraas E. Genome-Wide Association between Branch Point Properties and Alternative Splicing. Meyer IM, editor. *PLoS Comput Biol*. 2010;6:e1001016.
- Cyr JL, Brown GD, Stroop J, Heinen CD. The predicted truncation from a cancer-associated variant of the *MSH2* initiation codon alters activity of the MSH2-MSH6 mismatch repair complex: BIOCHEMISTRY OF AN MSH2 INITIATION CODON VARIANT. *Mol Carcinog*. 2012;51:647–58.
- D’Andrea AD, Grompe M. The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer*. 2003;3:23–34.
- Dabrowski M, Bukowy-Bieryllo Z, Zietkiewicz E. Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biology*. 2015;12:950–8.
- Daguenet E, Dujardin G, Valcárcel J. The pathogenicity of splicing defects: mechanistic insights into pre- mRNA processing inform novel therapeutic approaches. *EMBO Rep*. 2015;16:1640–55.
- Datta A, Adjiri A, New L, Crouse GF, Jinks-Robertson S. Mitotic Crossovers between Diverged Sequences Are Regulated by Mismatch Repair Proteins in *Saccharomyces cerevisiae*. *MOL CELL BIOL*. 1996;16:9.
- Davy G, Rousselin A, Goardon N, Castéra L, Harter V, Legros A, et al. Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur J Hum Genet*. 2017;25:1147–54.
- de Feraudy Y, Ben Yaou R, Wahbi K, Stalens C, Stantzou A, Laugel V, et al. Very Low Residual Dystrophin Quantity Is Associated with Milder Dystrophinopathy. *Ann Neurol*. 2021;89:280–92.
- de la Grange P. Altérations de l’épissage et maladies rares. *Med Sci (Paris)*. 2016;32:1111–9.
- de la Hoya M, Soukariéh O, Lòpez-Perolio I, Vega A, Walker LC, van Ierland Y, et al. Combined genetic and splicing analysis of *BRCA1* c.[594-2A>C; 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms. *Hum Mol Genet*. 2016;25:2256–68.
- Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*. 2009;37:e67–e67.
- Desrichard A, Bidet Y, Uhrhammer N, Bignon Y-J. CHEK2 contribution to hereditary breast cancer in non-BRCAfamilies. *Breast Cancer Res*. 2011;13:R119.
- Dhooge M, Baert-Desurmont S, Corsini C, Caron O, Andrieu N, Berthet P, et al. National recommendations of the French Genetics and Cancer Group - Unicancer on the modalities of multi-genes panel analyses in hereditary predispositions to tumors of the digestive tract. *European Journal of Medical Genetics*. 2020;63:104080.
- Di Blasi C. Mild muscular dystrophy due to a nonsense mutation in the *LAMA2* gene resulting in exon skipping. *Brain*. 2001;124:698–704.
- Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frébourg T, Tosi M, et al. 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: HUMAN MUTATION. *Human Mutation*. 2013;34:1547–57.
- Disset A, Bourgeois CF, Benmalek N, Claustres M, Stevenin J, Tuffery-Giraud S. An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. *Human Molecular Genetics*. 2006;15:999–1013.
- Domchek SM, Tang J, Stopfer J, Lilli DR, Hamel N, Tischkowitz M, et al. Biallelic Deleterious *BRCA1* Mutations in a Woman with Early-Onset Ovarian Cancer. *Cancer Discovery*. 2013;3:399–405.
- Domingo E, Niessen RC, Oliveira C, Alhopuro P, Moutinho C, Espín E, et al. BRAF-V600E is not involved in the colorectal tumorigenesis of HNPCC in patients with functional MLH1 and MSH2 genes. *Oncogene*. 2005;24:3995–8.
- Dominguez-Valentin M, Sampson JR, Seppälä TT, ten Broeke SW, Plazzer J-P, Nakken S, et al. Cancer risks by gene,

- age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. *Genet Med*. 2020;22:15–25.
- Duckett DR, Bronstein SM, Taya Y, Modrich P. hMutS α - and hMutS β -dependent phosphorylation of p53 in response to DNA methylator damage. *Proc Natl Acad Sci U S A*. 1999;96:12384–8.
- Duckett DR, Drummond JT, MURCHIEt AIH, REARDONt JT, SANCARt A, LILLEYt DMJ, et al. Human MutS α recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin- d(GpG) adduct. *Proc Natl Acad Sci USA*. 1996;5.
- Dufner-Almeida LG, do Carmo RT, Masotti C, Haddad LA. Understanding human DNA variants affecting pre-mRNA splicing in the NGS era. *Adv Genet*. 2019;103:39–90.
- Durkacz BW, Omidiji O, Gray DA, Shall S. (ADP-ribose)_n participates in DNA excision repair. *Nature*. 1980;283:593–6.
- Dutil J, Godoy L, Rivera-Lugo R, Arroyo N, Albino E, Negrón L, et al. No Evidence for the Pathogenicity of the *BRCA2* c.6937 + 594T>G Deep Intronic Variant: A Case–Control Analysis. *Genetic Testing and Molecular Biomarkers*. 2018;22:85–9.
- Dvinge H. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett*. 2018;592:2987–3006.
- Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. A Systematic Genetic Assessment of 1,433 Sequence Variants of Unknown Clinical Significance in the *BRCA1* and *BRCA2* Breast Cancer–Predisposition Genes. *The American Journal of Human Genetics*. 2007;81:873–83.
- Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al. Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. *N Engl J Med*. 2015;372:2243–57.
- Echevarría L, Aupy P, Goyenvalle A. Exon-skipping advances for Duchenne muscular dystrophy. *Human Molecular Genetics*. 2018;27:R163–72.
- Edelbrock MA, Kaliyaperumal S, Williams KJ. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2013;743–744:53–66.
- Ehrenstein MR. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *The EMBO Journal*. 1999;18:3484–90.
- Eichler EE. Genetic Variation, Comparative Genomics, and the Diagnosis of Disease. *N Engl J Med*. 2019;381:64–74.
- Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet*. 2017;18:599–612.
- Eisen J. A phylogenomic study of the MutS family of proteins. *Nucleic Acids Research*. 1998;26:4291–300.
- Eisinger F, Bressac B, Castaigne D, Cottu P-H, Lansac J, Lefranc J-P, et al. Identification et prise en charge des prédispositions héréditaires aux cancers du sein et de l’ovaire (mise à jour 2004). *Pathologie Biologie*. 2006;54:230–50.
- Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic Acids Research*. 2014;42:10681–97.
- Escobar-Hoyos L, Knorr K, Abdel-Wahab O. Aberrant RNA Splicing in Cancer. *Annu Rev Cancer Biol*. 2019;3:167–85.
- Evans BJ, Burke W, Jarvik GP. The FDA and Genomic Tests — Getting Regulation Right. *N Engl J Med*. 2015;372:2258–64.
- Faà V, Incani F, Meloni A, Corda D, Masala M, Baffico AM, et al. Characterization of a Disease-associated Mutation Affecting a Putative Splicing Regulatory Element in Intron 6b of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene. *Journal of Biological Chemistry*. 2009;284:30024–31.
- Fackenthal JD, Yoshimatsu T, Zhang B, de Garibay GR, Colombo M, De Vecchi G, et al. Naturally occurring *BRCA2* alternative mRNA splicing events in clinically relevant samples. *J Med Genet*. 2016;53:548–58.
- Fairbrother WG. Predictive Identification of Exonic Splicing Enhancers in Human Genes. *Science*. 2002;297:1007–

13.

Fairbrother WG, Holste D, Burge CB, Sharp PA. Single Nucleotide Polymorphism–Based Validation of Exonic Splicing Enhancers. Sean Eddy, editor. *PLoS Biol.* 2004;2:e268.

Fang WH, Modrich P. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. *J Biol Chem.* 1993;268:11838–44.

Feraudy Y, Ben Yaou R, Wahbi K, Stalens C, Stantzou A, Laugel V, et al. Very Low Residual Dystrophin Quantity Is Associated with Milder Dystrophinopathy. *Ann Neurol.* 2021;89:280–92.

Fica SM, Tuttle N, Novak T, Li N-S, Lu J, Koodathingal P, et al. RNA catalyses nuclear pre-mRNA splicing. *Nature.* 2013;503:229–34.

Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature.* 2014;513:120–3.

Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature.* 2018;562:217–22.

Fink D, Aebi S, Howell SB. The role of DNA mismatch repair in drug resistance. *Clin Cancer Res.* 1998;4:1–6.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell.* 1993;75:1027–38.

Flanigan KM, Dunn DM, von Niederhausern A, Soltanzadeh P, Howard MT, Sampson JB, et al. Nonsense mutation-associated Becker muscular dystrophy: interplay between exon definition and splicing regulatory elements within the DMD gene. *Hum Mutat.* 2011;32:299–308.

Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;32:557–63.

Fox-Walsh KL, Dou Y, Lam BJ, Hung S -p., Baldi PF, Hertel KJ. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proceedings of the National Academy of Sciences.* 2005;102:16176–81.

Fraile-Bethencourt E, Díez-Gómez B, Velásquez-Zapata V, Acedo A, Sanz DJ, Velasco EA. Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. Eng C, editor. *PLoS Genet.* 2017;13:e1006691.

Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, Acedo A, Velasco EA. Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays. *Front Genet.* 2018;9:188.

Fraile-Bethencourt E, Valenzuela-Palomo A, Díez-Gómez B, Caloca MJ, Gómez-Barrero S, Velasco EA. Minigene Splicing Assays Identify 12 Spliceogenic Variants of BRCA2 Exons 14 and 15. *Front Genet.* 2019;10:503.

Fredericks A, Cygan K, Brown B, Fairbrother W. RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules.* 2015;5:893–909.

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

Fuks F, Milner J, Kouzarides T. BRCA2 associates with acetyltransferase activity when bound to P/CAF. *Oncogene.* 1998;17:2531–4.

Fukui K. DNA Mismatch Repair in Eukaryotes and Bacteria. *Journal of Nucleic Acids.* 2010;2010:1–16.

Fukumaki Y, Ghosh PK, Benz EJ, Reddy VB, Lebowitz P, Forget BG, et al. Abnormally spliced messenger RNA in erythroid cells from patients with beta+ thalassemia and monkey cells expressing a cloned beta+ thalassaemic gene. *Cell.* 1982;28:585–93.

Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol.* 2010;653:249–57.

Gaildrat P, Krieger S, Di Giacomo D, Abdat J, Révillion F, Caputo S, et al. Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J Med Genet.* 2012;49:609–17.

Gallego CJ, Shirts BH, Bennette CS, Guzauskas G, Amendola LM, Horike-Pyne M, et al. Next-Generation Sequencing Panels for the Diagnosis of Colorectal Cancer and Polyposis Syndromes: A Cost-Effectiveness Analysis. *JCO.* 2015;33:2084–91.

- Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is γ UnAy. *Nucleic Acids Research*. 2008;36:2257–67.
- George CM, Alani E. Multiple cellular mechanisms prevent chromosomal rearrangements involving repetitive DNA. *Critical Reviews in Biochemistry and Molecular Biology*. 2012;47:297–313.
- Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, et al. Guidelines on Genetic Evaluation and Management of Lynch Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2014;147:502–26.
- Ginjaar IB, Kneppers AL, v d Meulen J-DM, Anderson LV, Bremmer-Bout M, van Deutekom JC, et al. Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. *Eur J Hum Genet*. 2000;8:793–6.
- Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ. Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to BRCA1 and BRCA2. *The American Journal of Human Genetics*. 2004;75:535–44.
- Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, et al. Comparative Analysis Identifies Exonic Splicing Regulatory Sequences—The Complex Definition of Enhancers and Silencers. *Molecular Cell*. 2006;22:769–81.
- Gorodetska I, Kozeretska I, Dubrovskaya A. *BRCA* Genes: The Role in Genome Stability, Cancer Stemness and Therapy Resistance. *J Cancer*. 2019;10:2109–27.
- Górski B, Jakubowska A, Huzarski T, Byrski T, Gronwald J, Grzybowska E, et al. A high proportion of founder *BRCA1* mutations in Polish breast cancer families: *BRCA1* Mutation in Polish Breast Cancer. *Int J Cancer*. 2004;110:683–6.
- Graff RE, Möller S, Passarelli MN, Witte JS, Skytthe A, Christensen K, et al. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. *Clinical Gastroenterology and Hepatology*. 2017;15:1256–64.
- Graham TA, Sottoriva A. Measuring cancer evolution from the genome: Measuring cancer evolution. *J Pathol*. 2017;241:183–91.
- Grandval P, Fabre AJ, Gaildrat P, Baert-Desurmont S, Buisine M-P, Ferrari A, et al. UMD-MLH1/MSH2/MSH6 databases: description and analysis of genetic variations in French Lynch syndrome families. *Database*. 2013;2013:bat036–bat036.
- Graveley BR. Sorting out the complexity of SR protein functions. *RNA*. 2000;6:1197–211.
- Grodecká L, Lockerová P, Ravčuková B, Buratti E, Baralle FE, Dušek L, et al. Exon First Nucleotide Mutations in Splicing: Evaluation of In Silico Prediction Tools. Spilianakis CB, editor. *PLoS ONE*. 2014a;9:e89570.
- Grodecká L, Kramárek M, Lockerová P, Kováčová T, Ravčuková B, Richterová R, et al. No major effect of the *CDH1* c.2440-6C>G mutation on splicing detected in last exon-specific splicing minigene assay: No Major Effect of *CDH1* C.2440-6C>G On Splicing. *Genes Chromosomes Cancer*. 2014b;53:798–801.
- Grover A, Houlden H, Baker M, Adamson J, Lewis J, Prihar G, et al. 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. *Journal of Biological Chemistry*. 1999;274:15134–43.
- Guillot D, Martin SA. Exploiting DNA mismatch repair deficiency as a therapeutic strategy. *Experimental Cell Research*. 2014;329:110–5.
- Gupta D, Heinen CD. The mismatch repair-dependent DNA damage response: Mechanisms and implications. *DNA Repair*. 2019;78:60–9.
- Han L, Vickers KC, Samuels DC, Guo Y. Alternative applications for distinct RNA sequencing strategies. *Brief Bioinform*. 2015;16:629–39.
- Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, et al. Hereditary Diffuse Gastric Cancer Syndrome: *CDH1* Mutations and Beyond. *JAMA Oncol*. 2015;1:23.
- Hartmann L. Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci*. 2008;Volume:3252.
- Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, et al. The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-derived Mutation. *Journal of Biological Chemistry*.

2001;276:14537–40.

Hawn MT, Umar A, Carethers JM, Marra G, Kunkel TA, Boland CR, et al. Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint. *Cancer Res.* 1995;55:3721–5.

Hendriks G, Morolli B, Calléja FMGR, Plomp A, Mesman RLS, Meijers M, et al. An Efficient Pipeline for the Generation and Functional Analysis of Human *BRCA2* Variants of Uncertain Significance. *Human Mutation.* 2014;n/a-n/a.

Hendriks YMC, Wagner A, Morreau H, Menko F, Stormorken A, Quehenberger F, et al. Cancer risk in hereditary nonpolyposis colorectal cancer due to *MSH6* mutations: impact on counseling and surveillance. *Gastroenterology.* 2004;127:17–25.

Heppner Goss K, Trzepacz C, Tuohy TMF, Groden J. Attenuated APC alleles produce functional protein from internal translation initiation. *Proceedings of the National Academy of Sciences.* 2002;99:8161–6.

Hernandez N, Keller W. Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. *Cell.* 1983;35:89–99.

Hernández-Imaz E, Martín Y, de Conti L, Melean G, Valero A, Baralle M, et al. Functional Analysis of Mutations in Exon 9 of *NF1* Reveals the Presence of Several Elements Regulating Splicing. Bardoni B, editor. *PLoS ONE.* 2015;10:e0141735.

Hickman MJ, Samson LD. Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proceedings of the National Academy of Sciences.* 1999;96:10764–9.

Høberg-Vetti H, Ognedal E, Buisson A, Vamre TBA, Ariansen S, Hoover JM, et al. The intronic *BRCA1* c.5407-25T>A variant causing partly skipping of exon 23—a likely pathogenic variant with reduced penetrance? *Eur J Hum Genet.* 2020;28:1078–86.

Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on *BRCA1* and *BRCA2* variants. *Hum Mutat.* 2012;33:1228–38.

Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pagès-Berhouet S, et al. Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat.* 2008;29:975–82.

Howard MT, Anderson CB, Fass U, Khatri S, Gesteland RF, Atkins JF, et al. Readthrough of dystrophin stop codon mutations induced by aminoglycosides. *Ann Neurol.* 2004;55:422–6.

Hsieh P, Yamane K. DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development.* 2008;129:391–407.

Hunter N, Borts RH. *Mlh1* is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes Dev.* 1997;11:1573–82.

Hustedt N, Durocher D. The control of DNA repair by the cell cycle. *Nat Cell Biol.* 2017;19:1–9.

Huyton T, Bates PA, Zhang X, Sternberg MJE, Freemont PS. The *BRCA1* C-terminal domain: structure and function. *Mutation Research/DNA Repair.* 2000;460:319–32.

Ibrahim A, Barnes DR, Dunlop J, Barrowdale D, Antoniou AC, Berg JN. Attenuated familial adenomatous polyposis manifests as autosomal dominant late-onset colorectal cancer. *Eur J Hum Genet.* 2014;22:1330–3.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431:931–45.

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet.* 2016;99:877–85.

Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176:535–548.e24.

Jagannathan S, Bradley RK. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res.* 2016;26:1639–50.

Jagmohan-Changur S, Poikonen T, Vilkki S, Launonen V, Wikman F, Orntoft TF, et al. *EXO1* variants occur commonly in normal population: evidence against a role in hereditary nonpolyposis colorectal cancer. *Cancer Res.* 2003;63:154–8.

- Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *The American Journal of Human Genetics*. 2016;98:1077–81.
- Jasin M, Rothstein R. Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harbor Perspectives in Biology*. 2013;5:a012740–a012740.
- Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and Familial Colon Cancer. *Gastroenterology*. 2010;138:2044–58.
- Jia X, Burugula BB, Chen V, Lemons RM, Jayakody S, Maksutova M, et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *The American Journal of Human Genetics*. 2021;108:163–75.
- Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*. 2006;7:335–46.
- Jordan B. Un Nobel pour la réparation de l'ADN: Prix Nobel de Chimie 2015 : Thomas Lindahl, Paul Modrich et Aziz Sankar : *Med Sci (Paris)*. 2016;32:117–9.
- Kadyrov FA, Dzantiev L, Constantin N, Modrich P. Endonucleolytic Function of MutL α in Human Mismatch Repair. *Cell*. 2006;126:297–308.
- Kadyrova LY, Kadyrov FA. Endonuclease activities of MutL α and its homologs in DNA mismatch repair. *DNA Repair*. 2016;38:42–9.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
- Karran P. Mechanisms of tolerance to DNA damaging therapeutic drugs. *Carcinogenesis*. 2001;22:1931–7.
- Kashima T, Manley JL. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet*. 2003;34:460–3.
- Kast K, Rhiem K, Wappenschmidt B, Hahnen E, Hauke J, Bluemcke B, et al. Prevalence of *BRCA1/2* germline mutations in 21 401 families with breast and ovarian cancer. *J Med Genet*. 2016;53:465–71.
- Kazazian HH, Boehm DC, Seltzer KW. ACMG recommendations for standards for interpretation of sequence variations. *Genet Med*. 2000;2:302–3.
- Ke S, Chasin LA. Context-dependent splicing regulation: Exon definition, co-occurring motif pairs and tissue specificity. *RNA Biology*. 2011;8:384–8.
- Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, Russo JJ, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*. 2011;21:1360–74.
- Keeling KM, Xue X, Gunn G, Bedwell DM. Therapeutics Based on Stop Codon Readthrough. *Annu Rev Genom Hum Genet*. 2014;15:371–94.
- Kerr ID, Cox HC, Moyes K, Evans B, Burdett BC, van Kan A, et al. Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *J Community Genet*. 2017;8:87–95.
- Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. 2019;16:713–32.
- Khan SG, Yamanegi K, Zheng Z-M, Boyle J, Imoto K, Oh K-S, et al. *XPC* branch-point sequence mutations disrupt U2 snRNP binding, resulting in abnormal pre-mRNA splicing in xeroderma pigmentosum patients. *Hum Mutat*. 2010;31:167–75.
- Kim Y-E, Ki C-S, Jang M-A. Challenges and Considerations in Sequence Variant Interpretation for Mendelian Disorders. *Ann Lab Med*. 2019;39:421–9.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Shendure J, Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. 2014;19.
- Kishore S, Khanna A, Stamm S. Rapid generation of splicing reporters with pSpliceExpress. *Gene*. 2008;427:104–10.
- Konarska MM, Sharp PA. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell*. 1987;49:763–74.
- Kornblihtt AR, Vibe-Pedersen K, Baralle FE. Human fibronectin: molecular cloning evidence for two mRNA species

differing by an internal segment coding for a structural domain. *The EMBO Journal*. 1984;3:221–6.

Kowalewski C, Bremer J, Gostynski A, Wertheim-Tysarowska K, Wozniak K, Bal J, et al. Amelioration of junctional epidermolysis bullosa due to exon skipping. *Br J Dermatol*. 2016;174:1375–9.

Krawczak M, Reiss J, Cooper David N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum Genet [Internet]*. 1992 [cited 2021 Jan 23];90. Available from: <http://link.springer.com/10.1007/BF00210743>

Krawczak M, Thomas NST, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat*. 2007;28:150–8.

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

Kunkel T, Loeb L. Fidelity of mammalian DNA polymerases. *Science*. 1981;213:765–7.

Kunkel TA, Erie DA. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet*. 2015;49:291–313.

Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci*. 2016;129:461–7.

Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol*. 2019;20:406–20.

Kuznetsov SG, Chang S, Sharan SK. Functional analysis of human *BRCA2* variants using a mouse embryonic stem cell-based assay. *Methods Mol Biol*. 2010;653:259–80.

Kuznetsov SG, Liu P, Sharan SK. Mouse embryonic stem cell-based functional assay to evaluate mutations in *BRCA2*. *Nat Med*. 2008;14:875–81.

Kwok C-T, Ward RL, Hawkins NJ, Hitchins MP. Detection of allelic imbalance in *MLH1* expression by pyrosequencing serves as a tool for the identification of germline defects in Lynch syndrome. *Familial Cancer*. 2010;9:345–56.

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*. 2018;46:D1062–7.

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

Lareau LF, Brenner SE. Regulation of Splicing Factors by Alternative Splicing and NMD Is Conserved between Kingdoms Yet Evolutionarily Flexible. *Molecular Biology and Evolution*. 2015;32:1072–9.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*. 2007;446:926–9.

Lattimore V, Pearson J, Morley-Bunker A, Investigators kConFab, Spurdle A, Robinson B, et al. Quantifying *BRCA1* and *BRCA2* mRNA Isoform Expression Levels in Single Cells. *IJMS*. 2019;20:693.

Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*. 2015;372:2509–20.

Lee A, Moon B-I, Kim TH. *BRCA1/BRCA2* Pathogenic Variant Breast Cancer: Treatment and Prevention Strategies. *Ann Lab Med*. 2020;40:114–21.

Lejeune F, Maquat LE. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current Opinion in Cell Biology*. 2005;17:309–15.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.

Leman R, Gaildrat P, Le Gac G, Ka C, Fichou Y, Audrezet M-P, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res*. 2018;46:7913–23.

Leman R, Tubeuf H, Raad S, Tournier I, Derambure C, Lanos R, et al. Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. *BMC Genomics*. 2020;21:86.

- Lewandowska MA. The missing puzzle piece: splicing mutations. *Int J Clin Exp Pathol*. 2013;6:2675–82.
- Li H, LaDuca H, Pesaran T, Chao EC, Dolinsky JS, Parsons M, et al. Classification of variants of uncertain significance in BRCA1 and BRCA2 using personal and family history of cancer from individuals in a large hereditary cancer multigene panel testing cohort. *Genet Med*. 2020;22:701–8.
- Li L, Biswas K, Habib LA, Kuznetsov SG, Hamel N, Kirchoff T, et al. Functional redundancy of exon 12 of *BRCA2* revealed by a comprehensive analysis of the c.6853A>G (p.I2285V) variant. *Hum Mutat*. 2009;30:1543–50.
- Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, et al. A unified mechanism for intron and exon definition and back-splicing. *Nature*. 2019;573:375–80.
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343:78–85.
- Ligtenberg MJL, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet*. 2009;41:112–7.
- Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proceedings of the National Academy of Sciences*. 2011;108:11093–8.
- Lim KR, Maruyama R, Yokota T. Eteplirsen in the treatment of Duchenne muscular dystrophy. *DDDT*. 2017;Volume11:533–45.
- Lim KH, Han Z, Jeon HY, Kach J, Jing E, Weyn-Vanhentenryck S, et al. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. *Nat Commun*. 2020;11:3501.
- Lin J, Tang X, Boulling A, Zou W, Masson E, Fichou Y, et al. First estimate of the scale of canonical 5' splice site GT>GC variants capable of generating wild-type transcripts. *Human Mutation*. 2019;40:1856–73.
- Lindblom A, Tannergård P, Werelius B, Nordenskjöld M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet*. 1993;5:279–82.
- Lindeboom RG, Vermeulen M, Lehner B, Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet*. 2019;51:1645–51.
- Littink KW, Pott J-WR, Collin RWJ, Kroes HY, Verheij JBG, Blokland EAW, et al. A Novel Nonsense Mutation in *CEP290* Induces Exon Skipping and Leads to a Relatively Mild Retinal Phenotype. *Invest Ophthalmol Vis Sci*. 2010;51:3646.
- Litton JK, Rugo HS, Ettl J, Hurvitz SA, Gonçalves A, Lee K-H, et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline *BRCA* Mutation. *N Engl J Med*. 2018;379:753–63.
- Liu D, Keijzers G, Rasmussen LJ. DNA mismatch repair and its many roles in eukaryotic cells. *Mutation Research/Reviews in Mutation Research*. 2017;773:174–87.
- Liu G, Yang D, Sun Y, Shmulevich I, Xue F, Sood AK, et al. Differing clinical impact of *BRCA1* and *BRCA2* mutations in serous ovarian cancer. *Pharmacogenomics*. 2012;13:1523–35.
- Liu H-X, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & Development*. 1998;12:1998–2012.
- Liu H-X, Cartegni L, Zhang MQ, Krainer AR. A mechanism for exon skipping caused by nonsense or missense mutations in *BRCA1* and other genes. *Nat Genet*. 2001;27:55–8.
- Livon D, Moretta J, Noguès C. Les femmes à haut risque de cancer du sein : quelle attitude ? *La Presse Médicale*. 2019;48:1092–100.
- Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The Vigorous Immune Microenvironment of Microsatellite Instable Colon Cancer Is Balanced by Multiple Counter-Inhibitory Checkpoints. *Cancer Discovery*. 2015;5:43–51.
- Lomonosov M. Stabilization of stalled DNA replication forks by the *BRCA2* breast cancer susceptibility protein. *Genes & Development*. 2003;17:3017–22.
- López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of

hereditary disease? FEBS Letters. 2005;579:1900–3.

Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. *Science*. 2017;355:1152–8.

Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet*. 2011;43:879–82.

Lovelock PK, Spurdle AB, Mok MT, Farrugia DJ, Lakhani SR, Healey S, et al. Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? *Breast Cancer Res*. 2007;9:R82.

Ludwig KK, Neuner J, Butler A, Geurts JL, Kong AL. Risk reduction and survival benefit of prophylactic surgery in BRCA mutation carriers, a systematic review. *The American Journal of Surgery*. 2016;212:660–9.

Lynch HT, Harris RE, Guirgis HA, Maloney K, Carmody LL, Lynch JF. Familial association of breast/ovarian carcinoma. *Cancer*. 1978;41:1543–9.

Lynch HT, Krush AJ. Carcinoma of the breast and ovary in three families. *Surg Gynecol Obstet*. 1971;133:644–8.

Lynch HT, Krush AJ, Lemon HM, Kaplan AR, Condit PT, Bottomley RH. Tumor variation in families with breast cancer. *JAMA*. 1972;222:1631–5.

Lynch HT, Snyder C, Casey MJ. Hereditary ovarian and breast cancer: what have we learned? *Annals of Oncology*. 2013;24:viii83–95.

Lynch HT. Hereditary Colorectal Cancer. *The New England Journal of Medicine*. 2003;14.

Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895–2015. *Nat Rev Cancer*. 2015;15:181–94.

Lynch PM, Lynch HT, Harris RE. Hereditary proximal colonic cancer. *Diseases of the Colon & Rectum*. 1977;20:661–8.

Maby P, Galon J, Latouche J-B. Frameshift mutations, neoantigens and tumor-specific CD8⁺ T cells in microsatellite unstable colorectal cancers. *Oncol Immunology*. 2016;5:e1115943.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science*. 2012;335:823–8.

Madanikia SA, Bergner A, Ye X, Blakeley JO. Increased risk of breast cancer in women with NF1. *Am J Med Genet*. 2012;158A:3056–60.

Mahdavi M, Nassiri M, Kooshyar MM, Vakili-Azghandi M, Avan A, Sandry R, et al. Hereditary breast cancer; Genetic penetrance and current status with BRCA. *J Cell Physiol*. 2018;jcp.27464.

Mallery DL. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *The EMBO Journal*. 2002;21:6755–62.

Manceau G, Karoui M, Charachon A, Delchier J-C, Sobhani I. Le syndrome HNPCC (hereditary non polyposis colorectal cancer) ou syndrome de Lynch : un syndrome en rapport avec une défaillance du système de réparation de l'ADN. *Bulletin du Cancer*. 2011;98:323–36.

Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, et al. Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. *Int J Cancer*. 2005;116:692–702.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–53.

Mansour MA. Ubiquitination: Friend and foe in cancer. *The International Journal of Biochemistry & Cell Biology*. 2018;101:80–93.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 2008;18:1509–17.

Martin SA, McCabe N, Mullarkey M, Cummins R, Burgess DJ, Nakabeppu Y, et al. DNA Polymerases as Potential Therapeutic Targets for Cancers Deficient in the DNA Mismatch Repair Proteins MSH2 or MLH1. *Cancer Cell*. 2010;17:235–48.

Martin SA, McCarthy A, Barber LJ, Burgess DJ, Parry S, Lord CJ, et al. Methotrexate induces oxidative DNA damage

and is selectively lethal to tumour cells with defects in the DNA mismatch repair gene *MSH2*. *EMBO Mol Med*. 2009;1:323–37.

Martinez JS, Baldeyron C, Carreira A. Molding BRCA2 function through its interacting partners. *Cell Cycle*. 2015;14:3389–95.

Martinez-Contreras R, Cloutier P, Shkreta L, Fiset J-F, Revil T, Chabot B. hnRNP proteins and splicing control. *Adv Exp Med Biol*. 2007;623:123–47.

Martinez-Outschoorn UE, Balliet R, Lin Z, Whitaker-Menezes D, Birbe RC, Bombonati A, et al. BRCA1 mutations drive oxidative stress and glycolysis in the tumor microenvironment: Implications for breast cancer prevention with antioxidant therapies. *Cell Cycle*. 2012;11:4402–13.

Masciari S, Dillon DA, Rath M, Robson M, Weitzel JN, Balmana J, et al. Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. *Breast Cancer Res Treat*. 2012;133:1125–30.

Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*. 2014;15:108–21.

Matlin AJ, Clark F, Smith CWJ. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*. 2005;6:386–98.

Mátyás G, Giunta C, Steinmann B, Hossle JP, Hellwig R. Quantification of single nucleotide polymorphisms: A novel method that combines primer extension assay and capillary electrophoresis: SNP Quantitation Via SNUPE and LIF-CE. *Hum Mutat*. 2002;19:58–68.

Mazoyer S, Dunning AM, Serova O, Dearden J, Puget N, Healey CS, et al. A polymorphic stop codon in BRCA2. *Nat Genet*. 1996;14:253–4.

Mazoyer S, Puget N, Perrin-Vidoz L, Lynch HT, Serova-Sinilnikova OM, Lenoir GM. A BRCA1 Nonsense Mutation Causes Exon Skipping. *The American Journal of Human Genetics*. 1998;62:713–5.

Mazurek A, Berardini M, Fishel R. Activation of Human MutS Homologs by 8-Oxo-guanine DNA Damage. *Journal of Biological Chemistry*. 2002;277:8260–6.

McGlincy NJ, Smith CWJ. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences*. 2008;33:385–93.

Medina-Arana V, Delgado L, Bravo A, Martín J, Fernández-Peralta AM, González-Aguilera JJ. Tumor spectrum in lynch syndrome, DNA mismatch repair system and endogenous carcinogens. *J Surg Oncol*. 2012;106:10–6.

Meeks HD, Song H, Michailidou K, Bolla MK, Dennis J, Wang Q, et al. BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *JNCIJ*. 2016;108:djv315.

Mehta A, Haber JE. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*. 2014;6:a016428–a016428.

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

Menzel T, Nähse-Kumpf V, Kousholt AN, Klein DK, Lund-Andersen C, Lees M, et al. A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance. *EMBO Rep*. 2011;12:705–12.

Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol*. 2012;30:99–104.

Mercer TR, Clark MB, Andersen SB, Brunck ME, Haerty W, Crawford J, et al. Genome-wide discovery of human splicing branchpoints. *Genome Res*. 2015;25:290–303.

Mercuri E, Muntoni F, Osorio AN, Tulinius M, Buccella F, Morgenroth LP, et al. Safety and effectiveness of ataluren: comparison of results from the STRIDE Registry and CINRG DMD Natural History Study. *Journal of Comparative Effectiveness Research*. 2020;9:341–60.

Mesman RLS, Calléja FMGR, de la Hoya M, Devilee P, van Asperen CJ, Vrieling H, et al. Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. *Genet Med*. 2020;22:1355–65.

Mesman RLS, Calléja FMGR, Hendriks G, Morolli B, Misovic B, Devilee P, et al. The functional impact of variants of uncertain significance in BRCA2. *Genet Med*. 2019;21:293–302.

- Miki Y, Swensen J, Shattuck-Eidens D, Futreal P, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994;266:66–71.
- Milner J, Ponder B, Hughes-Davies L, Seltmann M, Kouzarides T. Transcriptional activation functions in BRCA2. *Nature*. 1997;386:772–3.
- Min J, Choi ES, Hwang K, Kim J, Sampath S, Venkitaraman AR, et al. The Breast Cancer Susceptibility Gene BRCA2 Is Required for the Maintenance of Telomere Homeostasis*. *Journal of Biological Chemistry*. 2012;287:5091–101.
- Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, et al. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci USA*. 2015;112:E5189–98.
- Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, et al. Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med*. 2016;375:2154–64.
- Miyaki M, Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet*. 1997;17:271–2.
- Modrich P. Mechanisms in Eukaryotic Mismatch Repair. *Journal of Biological Chemistry*. 2006;281:30305–9.
- Moghadasi S, Meeks HD, Vreeswijk MP, Janssen LA, Borg Å, Ehrencrona H, et al. The BRCA1 c. 5096G>A p.Arg1699Gln (R1699Q) intermediate risk variant: breast and ovarian cancer risk estimation and recommendations for clinical management from the ENIGMA consortium. *J Med Genet*. 2018;55:15–20.
- Mojas N, Lopes M, Jiricny J. Mismatch repair-dependent processing of methylation damage gives rise to persistent single-stranded gaps in newly replicated DNA. *Genes & Development*. 2007;21:3342–55.
- Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, et al. Cancer risk and survival in *path_MMR* carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut*. 2018;67:1306–16.
- Montalban G, Bonache S, Moles-Fernández A, Gisbert-Beamud A, Tenés A, Bach V, et al. Screening of BRCA1/2 deep intronic regions by targeted gene sequencing identifies the first germline BRCA1 variant causing pseudoexon activation in a patient with breast/ovarian cancer. *J Med Genet*. 2019;56:63–74.
- Monteiro ANA, August A, Hanafusa H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. *Proceedings of the National Academy of Sciences*. 1996;93:13595–9.
- Monteiro AN, Bouwman P, Kousholt AN, Eccles DM, Millot GA, Masson J-Y, et al. Variants of uncertain clinical significance in hereditary breast and ovarian cancer genes: best practices in functional analysis for clinical annotation. *J Med Genet*. 2020;57:509–18.
- Morak M, Schaefer K, Steinke-Lange V, Koehler U, Keinath S, Massdorf T, et al. Full-length transcript amplification and sequencing as universal method to test mRNA integrity and biallelic expression in mismatch repair genes. *Eur J Hum Genet*. 2019;27:1808–20.
- Moretta J, Berthet P, Bonadona V, Caron O, Cohen-Haguenauer O, Colas C, et al. Recommandations françaises du Groupe Génétique et Cancer pour l'analyse en panel de gènes dans les prédispositions héréditaires au cancer du sein ou de l'ovaire. *Bulletin du Cancer*. 2018;105:907–17.
- Mottagui-Tabar S, A. Isaksson L. The influence of the 5' codon context on translation termination in *Bacillus subtilis* and *Escherichia coli* is similar but different from *Salmonella typhimurium*. *Gene*. 1998;212:189–96.
- Mount SM. A catalogue of splice junction sequences. *Nucl Acids Res*. 1982;10:459–72.
- Muir EG, Bell AJY, Barlow KA. Multiple primary carcinomata of the colon, duodenum, and larynx associated with kerato-acanthomata of the face. *Br J Surg*. 1967;54:191–5.
- Nadeau G, Boufaied N, Moisan A, Lemieux KM, Cayanan C, Monteiro ANA, et al. BRCA1 can stimulate gene transcription by a unique mechanism. *EMBO Rep*. 2000;1:260–5.
- Nagai. Differences in expression between transcripts using alternative promoters of hMLH1 gene and their correlation with microsatellite instability. *Oncol Rep* [Internet]. 2009 [cited 2021 Feb 3]; Available from: <http://www.spandidos-publications.com/or/22/2/265>
- Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, et al. Mismatch Repair Gene **PMS2** : Disease-Causing Germline Mutations Are Frequent in Patients Whose Tumors Stain Negative for PMS2 Protein, but Paralogous Genes Obscure Mutation Detection and Interpretation. *Cancer Res*. 2004;64:4721–7.

- Naruse H, Ikawa N, Yamaguchi K, Nakamura Y, Arai M, Ishioka C, et al. Determination of splice-site mutations in Lynch syndrome (hereditary non-polyposis colorectal cancer) patients using functional splicing assay. *Familial Cancer*. 2009;8:509–17.
- Nazari I, Tayara H, Chong KT. Branch Point Selection in RNA Splicing Using Deep Learning. *IEEE Access*. 2019;7:1800–7.
- Nené NR, Reisel D, Leimbach A, Franchi D, Jones A, Evans I, et al. Association between the cervicovaginal microbiome, BRCA1 mutation status, and risk of ovarian cancer: a case-control study. *The Lancet Oncology*. 2019;20:1171–82.
- Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. *Genome Research*. 2001;11:863–74.
- Ngeow J, Sesock K, Eng C. Breast cancer risk and clinical implications for germline PTEN mutation carriers. *Breast Cancer Res Treat*. 2017;165:1–8.
- Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*. 2007;21:708–18.
- Nicolaides NC, Papadopoulos N, Liu B, Weit Y-F, Cartert KC, Rubent SM, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. 1994;371:6.
- Nilsen TW. RNA-RNA interactions in the spliceosome: Unraveling the ties that bind. *Cell*. 1994;78:1–4.
- Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010;463:457–63.
- Nix P, Mundt E, Coffee B, Goossen E, Warf BM, Brown K, et al. Interpretation of BRCA2 Splicing Variants: A Case Series of Challenging Variant Interpretations and the Importance of Functional RNA Analysis. *Familial Cancer* [Internet]. 2021 [cited 2021 Feb 3]; Available from: <http://link.springer.com/10.1007/s10689-020-00224-y>
- O'Brien V, Brown R. Signalling cell cycle arrest and cell death through the MMR System. *Carcinogenesis*. 2006;27:682–92.
- Oh JH, Kim YJ, Moon S, Nam H-Y, Jeon J-P, Ho Lee J, et al. Genotype instability during long-term subculture of lymphoblastoid cell lines. *J Hum Genet*. 2013;58:16–20.
- Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. Structural basis for recruitment of BRCA2 by PALB2. *EMBO Rep*. 2009;10:990–6.
- Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz H-J, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *The Lancet Oncology*. 2017;18:1182–91.
- Overman MJ, Lonardi S, Wong KYM, Lenz H-J, Gelsomino F, Aglietta M, et al. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *JCO*. 2018;36:773–9.
- Pacho F, Zambruno G, Calabresi V, Kiritsi D, Schneider H. Efficiency of translation termination in humans is highly dependent upon nucleotides in the neighbourhood of a (premature) termination codon. *Journal of Medical Genetics*. 2011;48:640–4.
- Padgett RA, Mount SM, Steitz JA, Sharp PA. Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell*. 1983;35:101–7.
- Pagani F. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Human Molecular Genetics*. 2003;12:1111–20.
- Pagani F, Raponi M, Baralle FE. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences*. 2005;102:6368–72.
- Paggi JM, Bejerano G. A sequence-based, deep learning model accurately predicts RNA splicing branchpoints. *RNA*. 2018;24:1647–58.
- Pagin A, Zerimech F, Leclerc J, Wacrenier A, Lejeune S, Descarpentries C, et al. Evaluation of a new panel of six mononucleotide repeat markers for the detection of DNA mismatch repair-deficient tumours. *Br J Cancer*. 2013;108:2079–87.
- Palmirotta R, Veri MC, Curia MC, Aceto G, D'Amico F, Esposito DL, et al. Transcripts with splicings of exons 15 and

16 of the hMLH1 gene in normal lymphocytes: implications in RNA-based mutation screening of hereditary non-polyposis colorectal cancer. *European Journal of Cancer*. 1998;34:927–30.

Paluch-Shimon S, Cardoso F, Sessa C, Balmana J, Cardoso MJ, Gilbert F, et al. Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO Clinical Practice Guidelines for cancer prevention and screening. *Annals of Oncology*. 2016;27:v103–10.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40:1413–5.

Papadopoulos N, Nicolaides N, Wei Y, Ruben S, Carter K, Rosen C, et al. Mutation of a mutL homolog in hereditary colon cancer. *Science*. 1994;263:1625–9.

Park SA, Ahn SI, Gallo J-M. Tau mis-splicing in the pathogenesis of neurodegenerative disorders. *BMB Reports*. 2016;49:405–13.

Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, Feliubadaló L, et al. Large scale multifactorial likelihood quantitative analysis of *BRCA1* and *BRCA2* variants: An ENIGMA resource to support clinical variant classification. *Human Mutation*. 2019;40:1557–78.

Patel AA, Steitz JA. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*. 2003;4:960–70.

Paton DM. Nusinersen: antisense oligonucleotide to increase SMN protein production in spinal muscular atrophy. *Drugs Today*. 2017;53:327.

Pavlov YI, Mian IM, Kunkel TA. Evidence for Preferential Mismatch Repair of Lagging Strand DNA Replication Errors in Yeast. *Current Biology*. 2003;13:744–8.

Peasland A, Matheson E, Hall A, Irving J. Alternative splicing of hMLH1 in childhood acute lymphoblastic leukaemia and characterisation of the variably expressed $\Delta 9/10$ isoform as a dominant negative species. *Leukemia Research*. 2010;34:322–7.

Peltomäki P. Epigenetic mechanisms in the pathogenesis of Lynch syndrome: Epigenetic mechanisms in the pathogenesis of Lynch syndrome. *Clin Genet*. 2014;85:403–12.

Peltomäki P, Gao X, Mecklin JP. Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations. *Fam Cancer*. 2001;1:9–15.

Peltomäki P, Lothe RA, Aaltonen LA, Pylkkänen L, Nyström-Lahti M, Seruca R, et al. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res*. 1993;53:5853–5.

Peltz SW, Morsy M, Welch EM, Jacobson A. Ataluren as an Agent for Therapeutic Nonsense Suppression. *Annu Rev Med*. 2013;64:407–25.

Phuah S-Y, Looi L-M, Hassan N, Rhodes A, Dean S, Taib NAM, et al. Triple-negative breast cancer and PTEN (phosphatase and tensin homologue) loss are predictors of BRCA1 germline mutations in women with early-onset and familial breast cancer, but not in women with isolated late-onset breast cancer. *Breast Cancer Res*. 2012;14:R142.

Pineda JMB, Bradley RK. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev*. 2018;32:577–91.

Pinotti M, Caruso P, Canella A, Campioni M, Tagariello G, Castaman G, et al. Ribosome readthrough accounts for secreted full-length factor IX in hemophilia B patients with nonsense mutations. *Hum Mutat*. 2012;33:1373–6.

Pitout I, Flynn LL, Wilton SD, Fletcher S. Antisense-mediated splice intervention to treat human disease: the odyssey continues. *F1000Res*. 2019;8:710.

Plaschke J, Bulitta C, Saeger HD, Schackert HK. Quantitative Differences between Aberrant Transcripts Which Occur as Common Isoforms and due to Mutation-based Exon Skipping of the Mismatch Repair Gene hMLH1. *Clinical Chemistry and Laboratory Medicine [Internet]*. 1999 [cited 2021 Jan 23];37. Available from: <https://www.degruyter.com/view/j/cclm.1999.37.issue-9/cclm.1999.131/cclm.1999.131.xml>

Plazzer JP, Sijmons RH, Woods MO, Peltomäki P, Thompson B, Den Dunnen JT, et al. The InSiGHT database: utilizing 100 years of insights into Lynch Syndrome. *Familial Cancer*. 2013;12:175–80.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum*

- Mutat. 2008;29:1282–91.
- Popp MW-L, Maquat LE. Organizing Principles of Mammalian Nonsense-Mediated mRNA Decay. *Annu Rev Genet.* 2013;47:139–65.
- Pozzoli U, Sironi M. Silencers regulate both constitutive and alternative splicing events in mammals. *CMLS, Cell Mol Life Sci.* 2005;62:1579–604.
- Prakash R, Zhang Y, Feng W, Jasin M. Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. *Cold Spring Harb Perspect Biol.* 2015;7:a016600.
- Pujade-Lauraine E, Ledermann JA, Selle F, Gebski V, Penson RT, Oza AM, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *The Lancet Oncology.* 2017;18:1274–84.
- Pyne MT, Brothman AR, Ward B, Pruss D, Hendrickson BC, Scholl T. The BRCA2 genetic variant IVS7 + 2T-->G is a mutation. *J Hum Genet.* 2000;45:351–7.
- Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, et al. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet.* 2011;43:1104–7.
- Rahman MA, Nasrin F, Bhattacharjee S, Nandi S. Hallmarks of Splicing Defects in Cancer: Clinical Applications in the Era of Personalized Medicine. *Cancers.* 2020;12:1381.
- Rahman N. Realizing the promise of cancer predisposition genes. *Nature.* 2014;505:302–8.
- Rahner N, Friedrichs N, Wehner M, Steinke V, Aretz S, Friedl W, et al. Nine novel pathogenic germline mutations in *MLH1*, *MSH2*, *MSH6* and *PMS2* in families with Lynch syndrome. *Acta Oncologica.* 2007;46:763–9.
- Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA Synthesis in Mammalian Cells. Schibler U, editor. *PLoS Biol.* 2006;4:e309.
- Ramalho AS, Clarke LA, Sousa M, Felicio V, Barreto C, Lopes C, et al. Comparative ex vivo, in vitro and in silico analyses of a CFTR splicing mutation: Importance of functional studies to establish disease liability of mutations. *Journal of Cystic Fibrosis.* 2016;15:21–33.
- Raponi M, Baralle FE, Pagani F. Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: the case of CFTR exon 12. *Nucleic Acids Research.* 2006;35:606–13.
- Raponi M, Douglas AGL, Tammaro C, Wilson DI, Baralle D. Evolutionary Constraint Helps Unmask a Splicing Regulatory Region in BRCA1 Exon 11. Prokunina-Olsson L, editor. *PLoS ONE.* 2012;7:e37255.
- Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, et al. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in *BRCA1* exon 6. *Hum Mutat.* 2011;32:436–44.
- Rath A, Mishra A, Ferreira VD, Hu C, Omerza G, Kelly K, et al. Functional interrogation of Lynch syndrome-associated *MSH2* missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Human Mutation.* 2019;40:2044–56.
- Raud L, Ka C, Gourlaouen I, Callebaut I, Férec C, Le Gac G, et al. Functional analysis of novel *RHD* variants: splicing disruption is likely to be a common mechanism of variant D phenotype: SPLICING DEFECT AND VARIANT D PHENOTYPE. *Transfusion.* 2019;59:1367–75.
- Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, et al. Mutational spectrum in a worldwide study of 29,700 families with *BRCA1* or *BRCA2* mutations. *Human Mutation.* 2018;39:593–620.
- Reed R. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev.* 1996;6:215–20.
- Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* 1997;4:311–23.
- Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen — The Clinical Genome Resource. *N Engl J Med.* 2015;372:2235–42.
- Reilly NM, Yard BD, Pittman DL. Homologous Recombination-Mediated DNA Repair and Implications for Clinical Treatment of Repair Defective Cancers. *Methods Mol Biol.* 2019;1999:3–29.

- Reyes GX, Schmidt TT, Kolodner RD, Hombauer H. New insights into the mechanism of DNA mismatch repair. *Chromosoma*. 2015;124:443–62.
- Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008;10:294–300.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–24.
- Richman S. Deficient mismatch repair: Read all about it (Review). *International Journal of Oncology*. 2015;47:1189–202.
- Rizki A, Lundblad V. Defects in mismatch repair promote telomerase-independent proliferation. *Nature*. 2001;411:713–6.
- Roa BB, Boyd AA, Volcik K, Richards CS. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat Genet*. 1996;14:185–7.
- Robson M, Im S-A, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline *BRCA* Mutation. *N Engl J Med*. 2017;377:523–33.
- Rodnina MV, Korniy N, Klimova M, Karki P, Peng B-Z, Senyushkina T, et al. Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Research*. 2020;48:1056–67.
- Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Srivastava S, Jass JR, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: Meeting Highlights and Bethesda Guidelines. *JNCI Journal of the National Cancer Institute*. 1997;89:1758–62.
- Romero A, García-García F, López-Perolio I, Ruiz de Garibay G, García-Sáenz JA, Garre P, et al. BRCA1 Alternative splicing landscape in breast tissue samples. *BMC Cancer*. 2015;15:219.
- Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. *Cell*. 2015;163:698–711.
- Rosenthal ET, Bowles KR, Pruss D, van Kan A, Vail PJ, McElroy H, et al. Exceptions to the rule: Case studies in the prediction of pathogenicity for genetic variants in hereditary cancer genes. *Clin Genet*. 2015;88:533–41.
- Rospert S, Rakwalska M, Dubaquié Y. Polypeptide chain termination and stop codon readthrough on eukaryotic ribosomes. *Rev Physiol Biochem Pharmacol* [Internet]. 2005 [cited 2021 Feb 3]; Available from: <http://link.springer.com/10.1007/s10254-005-0039-7>
- Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer*. 2012;12:68–78.
- Ruffner H, Verma IM. BRCA1 is a cell cycle-regulated nuclear phosphoprotein. *Proceedings of the National Academy of Sciences*. 1997;94:7138–43.
- Ruijs MWG, Verhoef S, Rookus MA, Prunten R, van der Hout AH, Hogervorst FBL, et al. TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. *Journal of Medical Genetics*. 2010;47:421–8.
- Saha T, Rih JK, Rosen EM. BRCA1 down-regulates cellular levels of reactive oxygen species. *FEBS Letters*. 2009;583:1535–43.
- Saint-Martin C, Cauchois-Le Mière M, Rex E, Soukarieh O, Arnoux J, Buratti J, et al. Functional characterization of *ABCC8* variants of unknown significance based on bioinformatics predictions, splicing assays, and protein analyses: Benefits for the accurate diagnosis of congenital hyperinsulinism. *Human Mutation*. 2021;humu.24164.
- Sameer AS, Nissar S, Fatima K. Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *European Journal of Cancer Prevention*. 2014;23:246–57.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977;74:5463–7.
- Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, et al. Defective Mismatch Repair As a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer. *JCO*. 2010;28:3219–26.

- Schafer CM, Campbell NG, Cai G, Yu F, Makarov V, Yoon S, et al. Whole exome sequencing reveals minimal differences between cell line and whole blood derived DNA. *Genomics*. 2013;102:270–7.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*. 1995;270:467–70.
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell*. 2000;101:671–84.
- Schneider C, Will CL, Makarova OV, Makarov EM, Lührmann R. Human U4/U6.U5 and U4atac/U6atac.U5 Tri-snRNPs Exhibit Similar Protein Compositions. *MCB*. 2002;22:3219–29.
- Schrader CE, Edelman W, Kucherlapati R, Stavnezer J. Reduced Isotype Switching in Splenic B Cells from Mice Deficient in Mismatch Repair Enzymes. *Journal of Experimental Medicine*. 1999;190:323–30.
- Schueren F, Thoms S. Functional Translational Readthrough: A Systems Biology Perspective. Brosius J, editor. *PLoS Genet*. 2016;12:e1006196.
- Schwartz S, Hall E, Ast G. SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Research*. 2009;37:W189–92.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11:361–2.
- Schweingruber C, Rufener SC, Zünd D, Yamashita A, Mühlemann O. Nonsense-mediated mRNA decay — Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2013;1829:612–23.
- Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17:19–32.
- Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol*. 2019;20:698–714.
- Sedic M, Skibinski A, Brown N, Gallardo M, Mulligan P, Martinez P, et al. Haploinsufficiency for BRCA1 leads to cell-type-specific genomic instability and premature senescence. *Nat Commun*. 2015;6:7505.
- Senapathy P, Shapiro MB, Harris NL. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol*. 1990;183:252–78.
- Sevcik J, Falk M, Kleiblova P, Lhota F, Stefancikova L, Janatova M, et al. The BRCA1 alternative splicing variant $\Delta 14-15$ with an in-frame deletion of part of the regulatory serine-containing domain (SCD) impairs the DNA repair capacity in MCF-7 cells. *Cellular Signalling*. 2012;24:1023–30.
- Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*. 1987;15:7155–74.
- Sharma N, Sosnay PR, Ramalho AS, Douville C, Franca A, Gottschalk LB, et al. Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with In Silico Predictions. *Human Mutation*. 2014;35:1249–59.
- Sharp A, Pichert G, Lucassen A, Eccles D. RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. *Hum Mutat*. 2004;24:272–272.
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. *Nature*. 2017;550:345–53.
- Shepard PJ, Hertel KJ. Conserved RNA secondary structures promote alternative splicing. *RNA*. 2008;14:1463–9.
- Shepard PJ, Hertel KJ. The SR protein family. *Genome Biol*. 2009;10:242.
- Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*. 2001;29:308–11.
- Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research*. 2006;34:3955–67.
- Shevelev IV, Hübscher U. The 3'–5' exonucleases. *Nat Rev Mol Cell Biol*. 2002;3:364–76.
- Shi Y. The Spliceosome: A Protein-Directed Metalloribozyme. *Journal of Molecular Biology*. 2017;429:2640–53.
- Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol*. 2017;18:655–70.

- Shiga N, Takeshima Y, Sakamoto H, Inoue K, Yokota Y, Yokoyama M, et al. Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J Clin Invest*. 1997;100:2204–10.
- Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C, et al. *BRCA2* Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. *Cancer Res*. 2017;77:2789–99.
- Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. *Nat Rev Genet*. 2016;17:407–21.
- Signal B, Gloss BS, Dinger ME, Mercer TR. Machine learning annotation of human branchpoints. Hancock J, editor. *Bioinformatics*. 2018;34:920–7.
- Singer-Sam J. Quantitation of specific transcripts by RT-PCR SNUPE assay. *Genome Research*. 1994;3:S48–50.
- Singh AK, Yu X. Tissue-Specific Carcinogens as Soil to Seed *BRCA1/2*-Mutant Hereditary Cancers. *Trends in Cancer*. 2020;6:559–68.
- Singh NN, Singh RN, Androphy EJ. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Research*. 2006;35:371–89.
- Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Human Molecular Genetics*. 2006;15:2490–508.
- Smyrk TC, Watson P, Kaul K, Lynch HT. Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma. *Cancer*. 2001;91:2417–22.
- Sonnenblick A, de Azambuja E, Azim HA, Piccart M. An update on PARP inhibitors—moving to the adjuvant setting. *Nat Rev Clin Oncol*. 2015;12:27–41.
- Sopik V, Akbari MR, Narod SA. Genetic testing for *RAD51C* mutations: in the clinic and community: *RAD51C* and ovarian cancer. *Clin Genet*. 2015;88:303–12.
- Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. Aretz S, editor. *PLoS Genet*. 2016;12:e1005756.
- Southey MC, Tesoriero A, Young MA, Holloway AJ, Jenkins MA, Whitty J, et al. A specific GFP expression assay, penetrance estimate, and histological assessment for a putative splice site mutation in *BRCA1*. *Hum Mutat*. 2003;22:86–91.
- Spurdle AB, Whiley PJ, Thompson B, Feng B, Healey S, Brown MA, et al. *BRCA1* R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. *J Med Genet*. 2012;49:525–32.
- Spurdle AB, Couch FJ, Hogervorst FBL, Radice P, Sinilnikova OM, for the IARC Unclassified Genetic Variants Working Group. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat*. 2008;29:1304–13.
- Spurdle AB, Couch FJ, Parsons MT, McGuffog L, Barrowdale D, Bolla MK, et al. Refined histopathological predictors of *BRCA1* and *BRCA2* mutation status: a large-scale analysis of breast cancer characteristics from the BCAC, CIMBA, and ENIGMA consortia. *Breast Cancer Res*. 2014;16:3419.
- Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL, et al. 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*. 2012;33:2–7.
- Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. *The American Journal of Human Genetics*. 2017;101:315–25.
- Stauffer S, Biswas K, Sharan SK. Bypass of premature stop codons and generation of functional *BRCA2* by exon skipping. *J Hum Genet*. 2020;65:805–9.
- Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. 2015;16:133–45.
- Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136:665–77.

- Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Research*. 2011;21:1563–71.
- Stevens M, Oltean S. Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing. *Front Genet*. 2019;10:804.
- Su SS, Lahue RS, Au KG, Modrich P. Mismatch specificity of methyl-directed DNA mismatch correction in vitro. *Journal of Biological Chemistry*. 1988;263:6829–35.
- Su SS, Modrich P. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proceedings of the National Academy of Sciences*. 1986;83:5057–61.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526:75–81.
- Supek F, Lehner B, Lindeboom RGH. To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. *Trends in Genetics*. 2020;S0168952520303073.
- Swann PF, Waters TR, Moulton DC, Xu Y-Z, Zheng Q, Edwards M, et al. Role of Postreplicative DNA Mismatch Repair in the Cytotoxic Action of Thioguanine. *Science*. 1996;273:1109–11.
- Szabo C, Masiello A, Ryan JF, Brody LC. The breast cancer information core: database design, structure, and scope. *Hum Mutat*. 2000;16:123–31.
- Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat*. 2008;29:1342–54.
- Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, Evolutionarily Unlikely Missense Substitutions in ATM Confer Increased Risk of Breast Cancer. *The American Journal of Human Genetics*. 2009;85:427–46.
- ten Broeke SW, van Bavel TC, Jansen AML, Gómez-García E, Hes FJ, van Hest LP, et al. Molecular Background of Colorectal Tumors From Patients With Lynch Syndrome Associated With Germline Variants in PMS2. *Gastroenterology*. 2018;155:844–51.
- Thanaraj TA. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Research*. 2001;29:2581–93.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487:330–7.
- Théry JC, Krieger S, Gaildrat P, Révillion F, Buisine M-P, Killian A, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet*. 2011;19:1052–8.
- Thibodeau S, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science*. 1993;260:816–9.
- Thibodeau SN, French AJ, Roche PC, Cunningham JM, Tester DJ, Lindor NM, et al. Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes. *Cancer Res*. 1996;56:4836–40.
- Thirthagiri E, Klarmann KD, Shukla AK, Southon E, Biswas K, Martin BK, et al. BRCA2 minor transcript lacking exons 4–7 supports viability in mice and may account for survival of humans with a pathogenic biallelic mutation. *Hum Mol Genet*. 2016;25:1934–45.
- Thompson BA, Martins A, Spurdle AB. A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays: A review of mismatch repair gene transcripts. *Clin Genet*. 2015;87:100–8.
- Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, et al. 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*. 2014;46:107–15.
- Thompson BA, Walters R, Parsons MT, Dumenil T, Drost M, Tiersma Y, et al. Contribution of mRNA Splicing to Mismatch Repair Gene Sequence Variant Interpretation. *Front Genet*. 2020;11:798.

- Thompson D, Easton DF, Goldgar DE. A Full-Likelihood Method for the Evaluation of Causality of Sequence Variants from Family Data. *The American Journal of Human Genetics*. 2003;73:652–5.
- Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat*. 2011;32:358–68.
- Torre D. Multiple sebaceous tumors. *Arch Dermatol*. 1968;98:549–51.
- Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, et al. A large fraction of unclassified variants of the mismatch repair genes *MLH1* and *MSH2* is associated with splicing defects. *Hum Mutat*. 2008;29:1412–24.
- Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, Yeats C, et al. The implications of alternative splicing in the ENCODE protein complement. *Proceedings of the National Academy of Sciences*. 2007;104:5495–500.
- Tress ML, Abascal F, Valencia A. Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends in Biochemical Sciences*. 2017;42:98–110.
- Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology*. 2002;122:211–9.
- Tsujii H, Iguchi Y, Furuya A, Kataoka A, Hatsuta H, Atsuta N, et al. Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO Mol Med*. 2013;5:221–34.
- Tubeuf H, Caputo SM, Sullivan T, Rondeaux J, Krieger S, Caux-Moncoutier V, et al. Calibration of Pathogenicity Due to Variant-Induced Leaky Splicing Defects by Using *BRCA2* Exon 3 as a Model System. *Cancer Res*. 2020a;80:3593–605.
- Tubeuf H, Charbonnier C, Soukarieh O, Blavier A, Lefebvre A, Dauchel H, et al. Large-scale comparative evaluation of user-friendly tools for predicting variant-induced alterations of splicing regulatory elements. *Human Mutation*. 2020b;41:1811–29.
- Tuffery-Giraud S, Saquet C, Thorel D, Disset A, Rivier F, Malcolm S, et al. Mutation spectrum leading to an attenuated phenotype in dystrophinopathies. *Eur J Hum Genet*. 2005;13:1254–60.
- Tung N, Domchek SM, Stadler Z, Nathanson KL, Couch F, Garber JE, et al. Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nat Rev Clin Oncol*. 2016;13:581–8.
- Turcot J, Despres JP, St Pierre F. Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases. *Dis Colon Rectum*. 1959;2:465–8.
- Umar A, Boland CR, Terdiman JP, Syngal S, Chapelle A d. l., Ruschoff J, et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *JNCI Journal of the National Cancer Institute*. 2004;96:261–8.
- Valle L, Vilar E, Tavtigian SV, Stoffel EM. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. *J Pathol*. 2019;247:574–88.
- van der Klift HM, Jansen AML, van der Steenstraten N, Bik EC, Tops CMJ, Devilee P, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med*. 2015;3:327–45.
- van Gent DC, Hoeijmakers JHJ, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet*. 2001;2:196–206.
- Van Heetvelde M, Van Bockstal M, Poppe B, Lambein K, Rosseel T, Atanesyan L, et al. Accurate detection and quantification of epigenetic and genetic second hits in *BRCA1* and *BRCA2* -associated hereditary breast and ovarian cancer reveals multiple co-acting second hits. *Cancer Letters*. 2018;425:125–33.
- van Oosten M, Stout GJ, Backendorf C, Rebel H, de Wind N, Darroudi F, et al. Mismatch repair protein Msh2 contributes to UVB-induced cell cycle arrest in epidermal and cultured mouse keratinocytes. *DNA Repair*. 2005;4:81–9.
- Vasen H, Watson P, Mecklin J, Lynch H. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC☆. *Gastroenterology*. 1999;116:1453–6.
- Vasen HFA, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. *Gut*. 2013;62:812–

23.

Vasen HFA, Johan G, Offerhaus A, Den Hartog Jager FCA, Menko FH, Nagengast FM, et al. The tumour spectrum in hereditary non-polyposis colorectal cancer: A study of 24 kindreds in the Netherlands. *Int J Cancer*. 1990;46:31–4.

Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, et al. Clinical analysis of *PMS2*: mutation detection and avoidance of pseudogenes. *Hum Mutat*. 2010;n/a-n/a.

Vears DF, Sénécal K, Borry P. Reporting practices for variants of uncertain significance from next generation sequencing technologies. *European Journal of Medical Genetics*. 2017;60:553–8.

Vo AT, Zhu F, Wu X, Yuan F, Gao Y, Gu L, et al. hMRE11 deficiency leads to microsatellite instability and defective DNA mismatch repair. *EMBO Rep*. 2005;6:438–44.

von Nicolai C, Ehlén Å, Martin C, Zhang X, Carreira A. A second DNA binding site in human BRCA2 promotes homologous recombination. *Nat Commun*. 2016;7:12813.

Wagner R, Meselson M. Repair tracts in mismatched DNA heteroduplexes. *Proceedings of the National Academy of Sciences*. 1976;73:4135–9.

Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*. 2009;136:701–18.

Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibir P, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med*. 2020;22:1005–14.

Walsh T, Casadei S, Munson KM, Eng M, Mandell JB, Gulsuner S, et al. CRISPR–Cas9/long-read sequencing approach to identify cryptic mutations in *BRCA1* and other tumour suppressor genes. *J Med Genet*. 2020;jmedgenet-2020-107320.

Wang G-S, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*. 2007;8:749–61.

Wang H, Elbein SC. Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol*. 2007;373:157–76.

Wang Q, Lasset C, Desseigne F, Frappaz D, Bergeron C, Navarro C, et al. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. *Cancer Res*. 1999a;59:294–7.

Wang Q, Zhang H, Fishel R, Greene MI. BRCA1 and cell signaling. *Oncogene*. 2000;19:6152–8.

Wang T-F, Kleckner N, Hunter N. Functional specificity of MutL homologs in yeast: Evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proceedings of the National Academy of Sciences*. 1999b;96:13914–9.

Wang Y, Qin J. MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proceedings of the National Academy of Sciences*. 2003;100:15387–92.

Wang Z, Burge CB. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*. 2008;14:802–13.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.

Wappenschmidt B, Becker AA, Hauke J, Weber U, Engert S, Köhler J, et al. Analysis of 30 Putative BRCA1 Splicing Mutations in Hereditary Breast and Ovarian Cancer Families Identifies Exonic Splice Site Mutations That Escape In Silico Prediction. Peterlongo P, editor. *PLoS ONE*. 2012;7:e50800.

Ward AJ, Cooper TA. The pathobiology of splicing: The pathobiology of splicing. *J Pathol*. 2010;220:152–63.

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171:737–8.

Wildenberg J, Meselson M. Mismatch repair in heteroduplex DNA. *Proceedings of the National Academy of Sciences*. 1975;72:2202–6.

Will CL, Lührmann R. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*. 2011;3:a003707–a003707.

- Will CL, Lührmann R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biological Chemistry* [Internet]. 2005 [cited 2021 Jan 23];386. Available from: <https://www.degruyter.com/view/j/bchm.2005.386.issue-8/bc.2005.084/bc.2005.084.xml>
- Win AK, Young JP, Lindor NM, Tucker KM, Ahnen DJ, Young GP, et al. Colorectal and Other Cancer Risks for Carriers and Noncarriers From Families With a DNA Mismatch Repair Gene Mutation: A Prospective Cohort Study. *JCO*. 2012;30:958–64.
- Wong AKC, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. RAD51 Interacts with the Evolutionarily Conserved BRC Motifs in the Human Breast Cancer Susceptibility Gene *brca2*. *Journal of Biological Chemistry*. 1997;272:31941–4.
- Wooster R, Neuhausen S, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. *Science*. 1994;265:2088–90.
- Wu K, Hinson SR, Ohashi A, Farrugia D, Wendt P, Tavtigian SV, et al. Functional Evaluation and Cancer Risk Assessment of *BRCA2* Unclassified Variants. *Cancer Res*. 2005;11.
- Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q, et al. Fanconi anemia is associated with a defect in the *BRCA2* partner *PALB2*. *Nat Genet*. 2007;39:159–61.
- Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, et al. Control of *BRCA2* Cellular and Clinical Functions by a Nuclear Partner, *PALB2*. *Molecular Cell*. 2006;22:719–29.
- Xia Y, Pao GM, Chen H-W, Verma IM, Hunter T. Enhancement of *BRCA1* E3 Ubiquitin Ligase Activity through Direct Interaction with the *BARD1* Protein. *Journal of Biological Chemistry*. 2003;278:5255–63.
- Xiong F, Li S. SF3b4: A Versatile Player in Eukaryotic Cells. *Front Cell Dev Biol*. 2020;8:14.
- Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science*. 2015;347:1254806–1254806.
- Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*. 2002;30:3754–66.
- Yan H. Snapshot of the allele-specific variation in human gene expression. *Methods Mol Biol*. 2005;311:31–8.
- Yang B, Damaschke N, Yao T, McCormick J, Wagner J, Jarrard D. Pyrosequencing for Accurate Imprinted Allele Expression Analysis: PYROSEQUENCING FOR ALLELE-SPECIFIC EXPRESSION. *J Cell Biochem*. 2015;116:1165–70.
- Yang H. *BRCA2* Function in DNA Binding and Recombination from a *BRCA2*-DSS1-ssDNA Structure. *Science*. 2002;297:1837–48.
- Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer Risks Associated With Germline *PALB2* Pathogenic Variants: An International Study of 524 Families. *JCO*. 2020;38:674–85.
- Yang Y. A simple two-step, “hit and fix” method to generate subtle mutations in BACs using short denatured PCR fragments. *Nucleic Acids Research*. 2003;31:80e–80.
- Yarosh DB, Rice M, Day RS, Foote RS, Mitra S. O6-Methylguanine-DNA methyltransferase in human cells. *Mutation Research/DNA Repair Reports*. 1984;131:27–36.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004;11:377–94.
- York SJ, Modrich P. Mismatch Repair-dependent Iterative Excision at Irreparable O6-Methylguanine Lesions in Human Nuclear Extracts*. *Journal of Biological Chemistry*. 2006;281:22674–83.
- Yoshioka K, Yoshioka Y, Hsieh P. ATR Kinase Activation Mediated by MutS α and MutL α in Response to Cytotoxic O6-Methylguanine Adducts. *Molecular Cell*. 2006;22:501–10.
- Young LC. DNA mismatch repair proteins promote apoptosis and suppress tumorigenesis in response to UVB irradiation: an in vivo study. *Carcinogenesis*. 2004;25:1821–7.
- Yu X. The BRCT Domain Is a Phospho-Protein Binding Domain. *Science*. 2003;302:639–42.
- Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, et al. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. *Gastroenterology*. 2015;149:604-613.e20.

- Yurgelun MB, Hampel H. Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention. American Society of Clinical Oncology Educational Book. 2018;101–9.
- Zaghlool A, Ameer A, Cavalier L, Feuk L. Splicing in the human brain. *Int Rev Neurobiol.* 2014;116:95–125.
- Zhang Q, Fan X, Wang Y, Sun M, Shao J, Guo D. BPP: a sequence-based algorithm for branch point prediction. Hancock J, editor. *Bioinformatics.* 2017;33:3166–72.
- Zhao W, Vaithiyalingam S, San Filippo J, Maranon DG, Jimenez-Sainz J, Fontenay GV, et al. Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry. *Molecular Cell.* 2015;59:176–87.
- Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, et al. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature.* 2011;477:179–84.