



HAL
open science

Recherche des bases moléculaires des phénotypes extrêmes de cancer par séquençage d'exome

Martin Chevarin

► **To cite this version:**

Martin Chevarin. Recherche des bases moléculaires des phénotypes extrêmes de cancer par séquençage d'exome. Cancer. 2019. hal-02457986

HAL Id: hal-02457986

<https://ephe.hal.science/hal-02457986>

Submitted on 5 Mar 2020

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.



École Pratique
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES

Sciences de la Vie et de la Terre

MEMOIRE

Présenté par

Martin CHEVARIN

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

Recherche des bases moléculaires des phénotypes

extrêmes de cancer par séquençage d'exome

Soutenu devant le jury suivant:

Dr Sophie GAD-LAPITEAU- Président

Pr Laurence OLIVIER - FAIVRE - Tuteur scientifique

Dr Stéphanie PLENCHETTE - COLAS - Tuteur pédagogique

Pr Mary CALLANAN - Rapporteur

Dr Jessica GOBBO - Examineur

Dr Juliette ALBUISSON - Examineur

Mémoire préparé sous la direction de :

Pr Laurence OLIVIER - FAIVRE

Centre de génétique

CHU Dijon Bourgogne - Hôpital François Mitterrand

et de

Dr Stéphanie PLENCHETTE - COLAS

EA 7269 Laboratoire d'Immunologie et Immunothérapie des Cancers (LIIC)

EPHE (Sciences de la Vie et de la Terre)

GRET : Génomes, Cancers et Thérapies (GCT)

Remerciements

Les travaux présentés dans ce mémoire ont été réalisés au sein de l'équipe de Génétique des Anomalies du Développement (GAD) faisant partie de l'unité INSERM 1231 Lipides, Nutrition et Cancer dirigée par le Professeur

Je tiens à remercier en premier lieux le Professeur Laurence Olivier - Faivre, directrice de l'équipe GAD pour m'avoir accueillie au sein de son équipe de recherche, m'avoir permis de réaliser ce diplôme et avoir assuré son encadrement scientifique ainsi que pour les connaissances qu'elle m'a transmises, le temps et l'attention qu'elle m'a consacré.

Pour l'honneur qu'ils me font en faisant partie de ce jury, je remercie profondément le Professeur Sophie Gad - Lapiteau qui a accepté de le présider, le Professeur Mary Callanan qui a bien voulu juger ce travail, et les Docteurs Juliette Albuissou et Jessica Gobbo qui l'ont l'examiné.

Je remercie l'ensemble des biologistes, ingénieurs, techniciens et étudiants de l'équipe GAD que je côtoie au quotidien. Merci tout particulièrement à Morgane pour les techniques de biologie cellulaire qu'elle a réalisés. Merci à Thibaud et Charlotte pour leur aide et les services qu'ils ont pu me rendre.

J'adresse également des remerciements particuliers aux professeurs Christel Thauvin et Christophe Philippe et aux docteurs Sophie Nambot et Laurence Jégo pour leur aide et leurs conseils dans la conception et l'interprétation de certaines expériences ainsi que leurs relectures de mon mémoire.

Je tiens à remercier également l'ensemble des cliniciens, biologistes et conseillers en génétique des équipes d'oncogénétique du CHU de Dijon et du centre Georges François Leclerc pour leurs contributions, et l'aide qu'ils ont pu m'apporter tout au long de ce diplôme.

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

Mes plus grands remerciements sont évidemment pour vous deux, Hélène et Mathieu, mes deux amours, qui me rappelez tous les jours ce qui est le plus important dans la vie.

Sommaire

Sommaire	1
Liste des figures	4
Liste des tableaux	5
Liste des abréviations	6
Avant-propos	8
Introduction	10
I. Bases moléculaires des cancers	10
I.1. Classes de gènes impliqués dans l'oncogenèse	11
I.1.a. Les proto-oncogènes	11
I.1.b. Les gènes suppresseurs de tumeur (gatekeepers)	12
I.1.c. Les gènes de réparation de l'ADN (caretakers)	13
I.2. Complexité génétique des tumeurs	15
I.2.a. Nombre de mutations	15
I.2.b. Hétérogénéité génétique	15
I.2.c. Voies de signalisation impliquées dans l'oncogenèse	16
II. Les prédispositions génétiques au cancer	19
II.1. Les gènes impliqués dans la prédisposition au cancer	20
II.2. Les principales prédispositions génétiques au cancer	23
II.3. Recherche des bases moléculaires des prédispositions génétiques au cancer	24
II.4. Conséquences de la découverte d'une variation pathogène dans un gène de prédisposition au cancer	24
III. Le séquençage haut débit	27
III.1. Aspects techniques du séquençage haut débit	27
III.1.a. Préparation des bibliothèques	28
III.1.b. Amplification clonale	28
III.1.c. Les différentes plateformes de séquençage haut débit	30
III.1.d. Traitement bioinformatique des données brutes	33
III.1.e. Interprétation des variations	35
III.2. Le séquençage de l'exome	36
III.2.a. Différentes stratégies d'analyse d'exome	36

III.2.b. Validation de la pathogénicité des variations.....	37
III.3. Apports scientifiques et médicaux du NGS/ES.....	38
III.3.a. Découverte de la variabilité du génome humain.....	38
III.3.b. Apport du NGS/ES dans les maladies Mendéliennes.....	39
III.3.c. Apport du NGS/ES en oncogénétique.....	39
Problématique et hypothèse de recherche.....	41
Matériel et méthodes.....	44
I. Extraction de l'ADN génomique.....	44
II. Dosage et contrôle qualité de l'ADN génomique.....	44
III. Extraction, dosage et contrôle qualité des ARN totaux.....	45
IV. Séquençage d'exome (ES).....	45
V. Analyse des données brutes de séquençage haut débit.....	45
VI. Séquençage haut débit ciblé.....	46
VII. Séquençage selon la méthode de Sanger.....	46
VIII. Recherche de perte d'hétérozygotie par analyse chromosomique sur puce à ADN (ACPA)	47
IX. Culture cellulaire et analyses fonctionnelles sur lignée cellulaires.....	47
IX.1. TP53I3.....	47
IX.2. ATR.....	49
IX.3. Anticorps.....	49
Résultats.....	50
I. Variations identifiées dans les cancers du sein masculins.....	54
I.1. ATR (Ataxia Telangiectasia and RAD3 related).....	54
I.1.a. Séquençage ciblé sur la cohorte de réplication.....	55
I.1.b. Identification d'une 3 ^{ème} variation d'ATR.....	56
I.1.c. Recherche de perte d'hétérozygotie dans la tumeur.....	57
I.1.d. Impact sur l'épissage.....	58
I.1.e. Expression d'ATR et phosphorylation de CHEK1.....	59
I.2. TP53I3 (Tumor Protein P53 Inducible protein 3).....	60
I.2.a. Recherche de réplication.....	61
I.2.b. Ségrégation familiale de la variation.....	61
I.2.c. Recherche de perte d'hétérozygotie.....	61

I.2.d. Expression protéique	62
I.2.e. Effet de la variation sur la mort cellulaire	63
I.3. DLC1 (Deleted in Liver Cancer 1)	64
I.3.a. Séquençage ciblé sur la cohorte de réplication	65
I.3.b. Impact sur l'épissage	66
II. Variations identifiées dans les autres présentations cliniques	67
II.1. Gènes connus comme étant responsables de prédisposition génétique au cancer	67
II.1.a. PALB2 (PARTner and Localizer of BRCA2)	67
II.1.b. BLM (BLM RecQ like helicase)	68
II.2. Nouveaux gènes candidats	70
II.2.a. BIK (BCL2-Intercating Killer)	70
II.2.b. DFFA (DNA Fragmentation Factor subunit Alpha)	71
II.2.c. FAT4 (FAT atypical cadherin 4)	72
II.2.d. SETD2 (SET domain containing 2, histone lysine methyltransferase)	73
Discussion	75
I. Cancers du sein masculin	75
II. Gènes connus comme étant responsables de prédisposition génétique au cancer	77
III. Gènes candidats et résultats négatifs	78
Conclusion générale et perspectives	81
Bibliographie	83

Liste des figures

Figure 1: Les caractéristiques des cellules tumorales selon D. Hanahan et RA. Weinberg	10
Figure 2: Distribution des mutations selon les différentes classes de gènes.	14
Figure 3: Principales voies de signalisation et processus cellulaires impliquées en cancérologie.	18
Figure 4: « Two hits model ».	21
Figure 5: Amplification clonales. A: PCR en émulsion. B: PCR en pont sur support solide.....	29
Figure 6: Séquençage par synthèse.....	31
Figure 7: Séquençage par ligation selon la technologie SOLiD.	32
Figure 8: Différentes étapes de l'analyse bioinformatique pour l'identification de variations génétiques.....	33
Figure 9: A: Proportion du risque familiale expliqué par les GPC connus et les SNP identifiés par GWAS pour trois cancers communs. B: Effet phénotypique et fréquence des allèles de prédisposition au cancer.....	42
Figure 10: Troisième variation du gène ATR (c.1A>G – p.Met1?)......	57
Figure 11: ACPA comparative ADN tumoral/ADN constitutionnel au locus d'ATR.....	58
Figure 12: Impact de la variation c.7762-2A>C sur l'épissage.....	59
Figure 13: A : Analyse de l'expression protéique d'ATR dans les LCL des patients 1 et 3. B: phosphorylation de CHK1 dans la LCL du patient 1.	60
Figure 14: Séquençage Sanger de la variation de TP53I3 chez le cas index (CI) et son fils.	62
Figure 15: Analyse de l'expression protéique de TP53I3 par western blot)	62
Figure 16: Effet de la variation de TP53I3 sur la mort cellulaire induite par un stress génétoxique	63
Figure 17: Etude de l'impact de la variation c.3167+5G>C sur l'épissage de <i>DLC1</i>	67
Figure 18: Confirmation des variations du gène <i>BLM</i> chez la patiente et ses deux filles par séquençage Sanger.	69
Figure 19: Séquençage ciblé de l'ARN messager de BIK..	70

Liste des tableaux

Tableau 1: Relation entre le score de Pred et la probabilité d'erreur de l'appel d'une base..	34
Tableau 2: Patients inclus et stratégie d'analyse d'exome	51
Tableau 3: Gènes candidats identifiés par séquençage d'exome	53
Tableau 4: Variations d' <i>ATR</i> identifiées dans la cohorte de réplication	56
Tableau 5: Résumé des 3 variations du gène <i>ATR</i> et du matériel disponible pour chaque patient	57
Tableau 6: Variations de <i>DLC1</i> identifiées dans la cohorte de réplication	66

Liste des abréviations

ACPA: Analyse Chromosomique sur Puce à ADN
ADN: Acide Desoxyribonucléique
BER: Base Excision Repair
BH: BCL2 Homology
CCR: Cancers ColoRectal
ccRCC: clear cell Renal Cell Carcinoma
CGFL: Centre anti cancer Georges François Leclerc
ChIp-Seq: Chromatin Immuno Precipitation - Sequencing
CNV: Copy Number Variation
COSMIC: Catalog of Somatic Mutation in Cancer
CSF: Cancer du Sein Féminin
CSM: Cancer du Sein Masculin
DFF: DNA Fragmentation Factor
ENCODE: ENCYclopedia Of DNA Elements
ES: Exome Sequencing
ExAC: Exome Aggregation Consortium
FFPE: Formol Fixed and Parafin Embended
GGC: Groupe Génétique et Cancer
GPC: Gène de Prédisposition au Cancer
GWAS : Genome Wide Association Studies
HDDR: Human DNA Damage Response Disorders
HES: Hématoxyline, Eosine, Safran
HGMD: Human Gene Mutation Database
HH: Voie de signalisation Hedghog
HNPCC: Hereditary Non-Polyposis Colorectal Cancer
HR : Homologous Recombination
HSF: Human Splice Finder
INCa: Institut National du Cancer
Indel: Insertions ou délétions
LCL: Lignées Cellulaires Lymphoblastiques
LOH: Loss Of Heterozygosity
MMR: MisMatchRepair
NER: Nucleotide Excision Repair
NGS: Next Generation Sequencing
NHEJ: NonHomologous End-Joining
NMD: Non-sens Mediated Decay
NSSI : Non Synonymous, Splice Site, Insertion/deletion
OMIM: Online Mendelian Inheritance in Man
ORL: OtoRhinoLaryngée
PAF: Polypose Adénomateuse Familiale
RCP: Réunion de Concertation Pluridisciplinaire
RHO-GAP: RHO-specific GTPase Accelerating Protein
ROS: Reactive Oxygen Species

SNV: Single Nucleotide Variation

SOLiD: Supported Oligonucleotide Ligation and Detection

WGS: Whole Genome Sequencing

Avant-propos

Le cancer demeure la première cause de mortalité en France malgré une diminution globale de l'incidence et du taux de mortalité attribuée à cette pathologie au cours des vingt dernières années. Pour l'année 2017, le nombre de nouveaux cas est estimé à 400 000 et le nombre de décès à 150 000 (INCa, 2015a).

Dans la majorité des cas, l'apparition d'un cancer est liée à l'accumulation d'altérations génétiques et épigénétiques acquises au cours de la vie par un petit nombre de cellules. Ces mutations sont dites somatiques. **Mais 5 à 10 % des cancers se présentent comme de véritables maladies héréditaires se transmettant selon un mode mendélien.** Ils sont dus à la présence d'altérations génétiques constitutionnelles présentes dès la naissance dans toutes les cellules de l'organisme et transmissibles à la descendance. Ces variations sont souvent responsables d'une **instabilité génétique** qui accélère l'apparition de mutations somatiques secondaires. Dans ces situations, le quota de mutations nécessaires à la transformation d'une cellule normale en cellule cancéreuse est atteint plus rapidement.

Les prédispositions génétiques au cancer se transmettent le plus souvent selon un **mode autosomique dominant avec une pénétrance incomplète et une expressivité variable.** Pour les personnes concernées, le risque de cancer peut être très élevé comparé au risque moyen de la population générale.

Il existe des formes de cancer que l'on peut qualifier de « **phénotypes extrêmes** ». Il s'agit :

- de formes **sporadiques particulièrement précoces**
- de formes **familiales avec un excès de cancers dans une même branche parentale**
- de patients présentant **plusieurs tumeurs primitives** ou des **tumeurs bilatérales** pour les organes doubles
- de l'association de **plusieurs cancers rares** chez un même patient ou au sein d'une même famille.

Devant ces présentations, une prédisposition génétique est très fortement suspectée.

Depuis 2010, la mise en place de techniques de séquençage haut débit dans les laboratoires d'oncogénétique a permis de réduire notablement les coûts et les délais de séquençage. Cela a rendu possible l'analyse simultanée de groupes de gènes (ou **panels**) connus pour être impliqués dans les formes mendéliennes de cancer. Cette technique est devenue la pratique courante pour identifier la mutation responsable d'une prédisposition au cancer. Cependant, les différents panels de gènes actuellement séquencés ne permettent d'établir une prédisposition génétique concernant un gène à forte pénétrance que dans **10% des cas** (INCa, 2016; Susswein et al., 2016). Cela montre que, malgré l'apport du séquençage haut débit,

l'analyse des gènes de prédisposition connus ne permet d'expliquer qu'une petite proportion des situations où une prédisposition génétique au cancer est fortement suspectée.

Parmi les différentes techniques de séquençage haut débit, le séquençage de l'exome (ES: Exome Sequencing) permet d'analyser 95 % des séquences codantes du génome d'un individu. Il s'est largement répandu dans le domaine de la génétique médicale, où il a notamment permis d'identifier de nombreux nouveaux gènes responsables de maladies mendéliennes, à partir d'un nombre limité de patients, voir d'un patient unique, en utilisant la stratégie de séquençage de trios dans les formes sporadiques (cas index et ses parents sains), avec un filtre sur les nouvelles mutations (Koboldt et al., 2013).

L'objectif de ce projet est de **mettre à profit les méthodologies issues de l'étude des maladies rares, et ainsi d'utiliser l'ES chez des patients présentant un cancer de phénotype extrême non expliqué par les approches diagnostiques classiques pour tenter d'en découvrir les bases moléculaires.**

Introduction

I. Bases moléculaires des cancers

L'apparition d'un cancer résulte de l'acquisition par certaines cellules de capacités particulières qui leur confèrent un avantage sélectif ou prolifératif et leur permettent d'échapper au contrôle homéostatique assurant le maintien de l'architecture et de l'activité normale des organes. Ces cellules sont alors capables de former des masses tumorales affectant le fonctionnement des organes au sein desquels elles se développent. A un stade avancé, elles peuvent acquérir un pouvoir migratoire et former des métastases dans d'autres organes. En 2000, puis 2011, D. Hanahan et RA. Weinberg ont défini les caractéristiques principales des cellules cancéreuses (Figure1) (Hanahan and Weinberg, 2000, 2011).

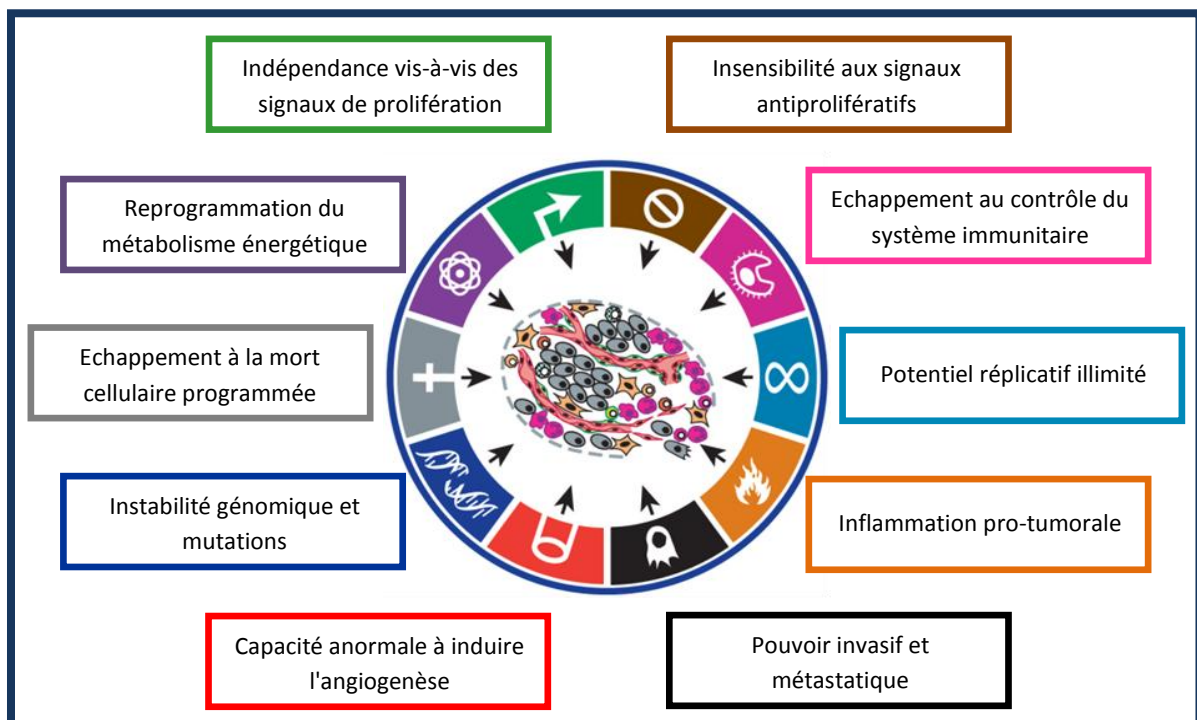


Figure 1: Les caractéristiques des cellules tumorales selon D. Hanahan et RA. Weinberg (Hanahan and Weinberg, 2000, 2011)

L'acquisition de ces nouvelles capacités est principalement due à l'accumulation d'altérations génétiques au sein du génome des cellules (Hanahan and Weinberg, 2011). Toutes sortes d'altérations génétiques sont présentes dans les tumeurs. La plupart sont des variations d'un seul nucléotide (SNV: *Single Nucleotide Variation*) ou des insertions ou délétions de quelques bases. On trouve également des événements de grande taille pouvant concerner plusieurs milliers de bases voir des bras chromosomiques entiers appelés variations du nombre

de copies (CNV: *Copy Number Variation*) ainsi que des translocations. Il est possible de dénombrer parfois plusieurs centaines de variations dans une tumeur (Greenman et al., 2007; Vogelstein et al., 2013; Wood et al., 2007). Cependant, toutes n'ont pas le même impact sur la destinée des cellules qu'elles affectent. On peut les classer en deux groupes: les mutations "*drivers*" qui confèrent un avantage sélectif aux cellules et interviennent directement dans les différents processus de transformation tumorale, et les mutations "*passengers*" qui ont un effet biologique neutre et ne sont que le reflet de l'instabilité génétique des cellules cancéreuses. La plupart des tumeurs contiennent entre deux et huit mutations "*drivers*" (Vogelstein et al., 2013). De plus, il existe une hétérogénéité cellulaire au sein d'une même tumeur avec la présence de différents clones de cellules cancéreuses. Certaines variations sont communes à tous les clones mais d'autres sont propres à chaque sous population (Gerlinger et al., 2012; Vogelstein et al., 2013).

I.1. Classes de gènes impliqués dans l'oncogenèse

De nombreux gènes impliqués dans le développement des cancers ont été identifiés. Le projet Cancer Gene Census porté par le consortium COSMIC (Catalog of Somatic Mutation in Cancer) liste 564 gènes dont il est établi que leur activité peut être liée au développement des cancers et que les mutations qui les affectent modifient leur activité en faveur de la transformation maligne. La majorité de ces gènes sont rapportés uniquement avec des mutations somatiques, qui n'affectent que les cellules tumorales ; 43 sont rapportés avec des variations constitutionnelles uniquement et 64 avec les deux types de variation. Ils peuvent être classés en trois catégories :

- les proto-oncogènes, dont l'activation va favoriser la prolifération et l'invasion cellulaire
- les gènes suppresseurs de tumeur qui vont au contraire bloquer ou freiner les processus conduisant à la transformation tumorale
- les gènes de réparation de l'ADN dont la fonction est d'assurer la stabilité du génome des cellules normales.

I.1.a. Les proto-oncogènes

La plupart des proto-oncogènes sont impliqués dans la transduction des différents signaux reçus par la cellule. Leur altération les rend hyperactifs favorisant la progression du cycle cellulaire, la croissance et la prolifération, et diminuant la mort cellulaire programmée.

L'activation des proto-oncogènes en oncogènes peut résulter d'une mutation ponctuelle avec un gain de fonction ou d'une amplification génique. La voie RAS/MAPK (Mitogen Activated Protein Kinase), anormalement activée dans de nombreux cancers, contient plusieurs exemples de proto-oncogènes. Elle est en général activée par la fixation d'un facteur de croissance sur un récepteur membranaire à activité tyrosine kinase comme le récepteur à l'EGF (EGFR: EpidermalGrowth Factor Receptor). Le signal est ensuite transmis à l'intérieur de la cellule par

une cascade de phosphorylation impliquant plusieurs autres protéines à activité kinase. Elle aboutit à l'expression de facteurs de transcription qui vont à leur tour induire l'expression d'un grand nombre de gènes ayant un rôle majeur dans l'initiation du cycle cellulaire. Des amplifications du gène de l'EGFR sont fréquemment retrouvées dans les cancers broncho-pulmonaires, de la sphère ORL (OtoRhinoLaryngée) et colorectaux. On trouve aussi fréquemment des mutations ponctuelles activatrices d'EGFR ou des kinases intracellulaires KRAS, NRAS ou BRAF dans différents types de tumeurs. Toutes ces altérations conduisent à une perte du contrôle et une suractivation anormale de la voie RAS/MAPK (Lièvre and Laurent-Puig, 2010).

Des translocations chromosomiques peuvent également être à l'origine de l'activation d'un proto-oncogène. Elles peuvent créer un gène de fusion avec une fonction anormale, comme la translocation récurrente t(9;22)(q34;q11) du chromosome de Philadelphie dans la leucémie myéloïde chronique, aboutissant à la création d'un gène de fusion entre les gènes *BCR* et *ABL*. Une translocation peut aussi juxtaposer une séquence de régulation active avec un proto-oncogène, augmentant ainsi son expression (O'Connor, 2008; Rowley, 2001).

Dans le cas d'altérations moléculaires conférant un gain de fonction ou une augmentation de l'expression, le mode de transmission est autosomique dominant. Cela signifie que la mutation constitutionnelle d'un seul allèle est suffisante pour avoir un effet cancérogène.

Les altérations gain de fonction sont le plus souvent des mutations faux-sens localisées à des positions bien précises appelées hot spots. Cette répartition des mutations est caractéristique des proto-oncogènes (Figure 2).

I.1.b. Les gènes suppresseurs de tumeur (gatekeepers)

Les gènes suppresseurs de tumeurs possèdent des propriétés inverses aux oncogènes puisqu'ils régulent négativement les voies de prolifération cellulaire et activent les points de contrôle du cycle cellulaire, permettant aux cellules de prévenir et corriger les erreurs de réplication de l'ADN avant d'entrer en mitose. Si les anomalies sont trop importantes, ils orientent alors la cellule vers un processus de mort programmée. L'inactivation de ces gènes peut abolir le contrôle négatif des voies de signalisation, favorisant la prolifération cellulaire et permettant aux cellules de continuer leur progression dans le cycle cellulaire malgré la présence d'anomalies de l'ADN qui devraient normalement conduire à son arrêt.

TP53 est le gène emblématique de cette catégorie. Il code pour le facteur de transcription P53 dont le rôle principal est d'activer l'expression de nombreux gènes impliqués dans le contrôle du cycle cellulaire, l'apoptose ou la réparation de l'ADN. En situation normale, P53 est constamment ubiquitinylé par MDM2, une ubiquitine ligase spécifique de P53. La quantité de P53 dans la cellule est relativement faible. Lorsque la cellule est soumise à un stress génotoxique, des modifications post traductionnelles de P53 vont déstabiliser le complexe P53-MDM2 et stabiliser P53. P53 va alors induire l'expression de nombreux gènes responsables de l'arrêt du cycle cellulaire et de la réparation des anomalies de l'ADN. Certains mécanismes de réparation sont directement contrôlés par P53. Une fois les réparations effectuées, le taux de

P53 revient à la normale et la cellule peut continuer sa progression dans le cycle cellulaire. Si les lésions sont trop importantes pour être réparées, P53 peut induire l'apoptose (Moll and Petrenko, 2003; Vogelstein et al., 2000). Le gène *TP53* est inactivé dans près de 50% des cancers humains. Des mutations constitutionnelles de ce gène sont responsables du syndrome de Li-Fraumeni (MIM #151623), une maladie rare prédisposant à de nombreux types de cancers avec un risque important pour les individus affectés de développer plusieurs cancers au cours de leur vie.

I.1.c. Les gènes de réparation de l'ADN (caretakers)

Contrairement aux proto-oncogènes et aux gènes suppresseurs de tumeurs, les gènes de réparation de l'ADN ne contrôlent pas directement le cycle cellulaire ou l'apoptose. Ils interviennent dans les différentes voies de réparation de l'ADN, indispensables au maintien de la stabilité du génome, en réparant les lésions de l'ADN qui peuvent survenir lors de la réplication ou sous l'effet de mutagènes externes ou endogènes. L'inactivation de ces gènes n'a pas pour conséquence directe une prolifération cellulaire accrue mais augmente de façon drastique l'instabilité de tout le génome en entraînant l'apparition de variations secondaires susceptibles d'activer les proto-oncogènes et d'inactiver les *gatekeepers* (Kinzler and Vogelstein, 1997). Il existe différents systèmes de réparation de l'ADN : le système MMR (*MisMatchRepair*), qui corrige les mésappariements de bases résultant d'erreurs de la réplication, les systèmes de réparation des cassures double brin par recombinaison homologue (HR: *Homologous Recombinaison*) ou non homologue (NHEJ: *NonHomologous End-Joining*) et le système de réparation par excision de base (BER : *Base Excision Repair*) ou de nucléotide (NER: *Nucleotide Excision Repair*) (Curtin, 2012).

Certains mécanismes de réparation sont fidèles alors que d'autres ne permettent qu'une réparation plus sommaire et laissent des erreurs. Cela peut être le cas lors de la réparation des cassures doubles brin. Le mécanisme de réparation par HR restaure fidèlement la séquence initiale. Elle n'est possible qu'en fin de phase S et en phase G2 du cycle cellulaire lorsque l'ADN à été répliqué et qu'une chromatide sœur est présente et peut servir de modèle pour recopier la séquence d'ADN à réparer. Dans les autres phases du cycle cellulaire, la réparation s'effectue par le système NHEJ qui rejoint simplement les deux extrémités de la cassure. Celui-ci change la séquence originelle et provoque une accumulation d'altérations génétiques (Wyman and Kanaar, 2006). En cas d'inactivation de certaines protéines impliquées dans la réparation par HR, comme BRCA1 ou BRCA2, les cassures d'ADN double brin ne peuvent être réparées que par le système NHEJ, ce qui conduit à un haut niveau d'instabilité génétique avec cassures chromosomiques, aneuploïdie et amplification du centrosome. Des mutations constitutionnelles des gènes BRCA1/2 conduisent à une augmentation du risque de cancer du sein et de l'ovaire (Kuchenbaecker et al., 2017).

Les mutations affectant les gènes suppresseurs de tumeur et les gènes de réparation sont le plus souvent des mutations perte de fonction. Il peut s'agir de SNV introduisant un codon stop prématuré ou de petites insertions/délétions entraînant un décalage du cadre de lecture. Les ARN messagers produits sont le plus souvent dégradés par le NMD (*Non-sens Mediated*

Decay) ou aboutissent à la production d'un polypeptide tronqué non fonctionnel. Plus rarement, on trouve des mutations faux-sens touchant un acide aminé particulièrement important pour la fonction du gène (Figure 2). Il est également possible de perdre un ou plusieurs gènes suppresseurs de tumeur par une grande délétion chromosomique. Par exemple, le bras court du chromosome 17 contenant le gène suppresseur de tumeur TP53 est fréquemment perdu dans les cancers colorectaux, du sein, de la prostate ou du cerveau (Fearon and Vogelstein, 1990).

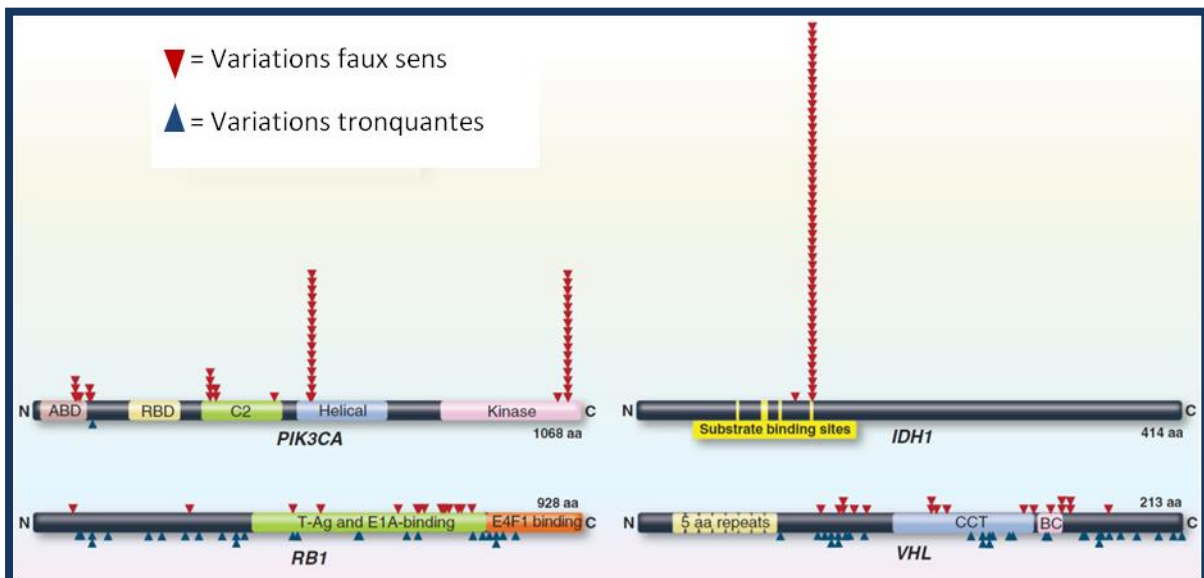


Figure 2: Distribution des mutations selon les différentes classes de gènes. Les proto-oncogènes *PIK3CA* et *IDH1* contiennent uniquement des mutations faux sens localisées à des positions récurrentes. Les mutations des gènes suppresseurs de tumeurs, *RB1* et *VHL* sont distribuées tout au long du gène et peuvent être des mutations tronquantes ou faux sens. Ces profils de mutation sont caractéristiques de la classe des gènes impliqués dans l'oncogenèse. Les variations sont issues de la base de données COSMIC (version 61). Pour *PIK3CA* et *IDH1*, les 50 premières ont été représentées. Pour *RB1* et *VHL*, toutes les variations sont représentées (Vogelstein et al., 2013).

Les mutations des gènes suppresseurs de tumeur et des gènes de réparation de l'ADN agissent le plus souvent selon un mode récessif, avec l'inactivation des deux allèles du gène pour que sa fonction soit perdue. Dans le cas d'une prédisposition génétique au cancer, il arrive fréquemment que le premier allèle soit porteur d'une mutation constitutionnelle et que le second allèle soit perdu par un événement somatique. Il s'agit du "**Two hits model**" proposé par AG. Knudson et DE. Comings (Comings, 1973; Knudson, 1971). Cependant, Il est maintenant admis que même une inactivation partielle de certains gènes suppresseurs de tumeur peut grandement contribuer à la tumorigénèse (Berger et al., 2011).

I.2. Complexité génétique des tumeurs

I.2.a. Nombre de mutations

Dans les tumeurs solides les plus communes (côlon, sein, ovaire, prostate), 33 à 66 gènes sont mutés en moyenne. Environ 95 % de ces variations sont des SNV (Vogelstein et al., 2013). Certaines tumeurs contiennent beaucoup plus de variations. Elles sont dues à l'exposition à des agents mutagènes comme la fumée de cigarette pour les cancers du poumon ou l'exposition aux UV pour les mélanomes, qui peuvent contenir jusqu'à 200 variations (Govindan et al., 2012; Wei et al., 2011). Un autre groupe de tumeurs hyper-mutées est celui des tumeurs liées à un défaut des systèmes de réparation de l'ADN comme les tumeurs du côlon avec défaut du système MMR dont le nombre médian de variations est supérieur à 700 (Cancer Genome Atlas Network, 2012). Des études récentes ont montré qu'un nombre élevé de variations sont également retrouvées dans les tumeurs avec altération génétique du domaine de correction des ADN polymérases POLE ou POLD1 (Cancer Genome Atlas Network, 2012; Palles et al., 2013). A l'inverse, les tumeurs pédiatriques et les hémopathies malignes contiennent beaucoup moins de variations ponctuelles (9,6 en moyenne) (Vogelstein et al., 2013).

Ce grand nombre de variations laisse envisager l'immense complexité du génome des cellules cancéreuses, mais cette apparence peut être trompeuse. En effet, il est très probable que la majorité des mutations soient des mutations *passengers*, qui n'octroient pas d'avantage sélectif ou prolifératif aux cellules et ne sont que le reflet de l'instabilité génétique des cellules cancéreuses. Le réel défi sur le plan thérapeutique est d'identifier quelles sont les quelques mutations *drivers* responsables de l'initiation et la progression du cancer. Les études épidémiologiques ont longtemps suggéré que les tumeurs solides contenaient cinq à huit mutations *drivers* (Armitage and Doll, 2004). Une étude plus récente a démontré que trois mutations *drivers* sont suffisantes au développement des cancers colorectaux et pulmonaires (Tomasetti et al., 2015).

I.2.b. Hétérogénéité génétique

Plusieurs études de séquençage pan-génomique des tumeurs ont mis en évidence une véritable hétérogénéité génétique intra-tumorale (Gerlinger et al., 2012; Navin et al., 2011; Nik-Zainal et al., 2012; Shah et al., 2009; Sottoriva et al., 2013; Xu et al., 2012; Yachida et al., 2010). Les tumeurs ou les métastases peuvent être considérées comme un ensemble dynamique de plusieurs populations de cellules transformées, ou en voie de transformation occupant des régions distinctes de la tumeur. Il est possible de réaliser un arbre phylogénétique retraçant l'apparition des mutations (Gerlinger et al. 2012). Les mutations *drivers* responsables de l'initiation et de la progression tumorale sont précoces et ubiquitaires. Elles représentent le tronc de l'arbre phylogénétique. Les mutations secondaires apparaissent dans des sous-clones cellulaires et représentent les branches de l'arbre. Elles ne sont pas essentielles au développement tumoral mais peuvent néanmoins favoriser l'extension métastatique et jouer un rôle important dans les phénomènes de résistance à la chimiothérapie.

La plupart des décès par cancer sont consécutifs au développement de métastases dans des organes comme le foie, le cerveau, les poumons ou les os, qui ne peuvent être traités chirurgicalement. Les patients peuvent présenter plusieurs métastases et l'hétérogénéité entre les différentes métastases d'un même patient peut être très importante. Il n'est pas rare d'observer dans une métastase une vingtaine de variations qui ne sont pas partagées par les autres lésions du même patient (Campbell et al., 2010; Yachida et al., 2010). Cette hétérogénéité inter-métastatique peut être expliquée par le fait que chaque métastase est formée par une cellule ou un groupe de cellules appartenant à des sous-populations clonales provenant de zones géographiquement distinctes de la tumeur primaire. Si cette hétérogénéité inter-métastase est importante, les différentes chimiothérapies ou traitements ciblés pouvant être envisagés n'auront un effet que sur une petite partie des lésions et le pronostic de survie sera défavorable.

Il existe également une hétérogénéité inter-patient, caractérisée par des parcours cliniques et des réponses thérapeutiques très différents entre des patients atteints d'un même type de cancer. Ces variations peuvent être dues à des facteurs génétiques constitutionnels de l'hôte qui influent sur la réponse au traitement (vitesse de métabolisation des drogues, perméabilité vasculaire) ou à des facteurs environnementaux. Cependant, la part la plus importante de cette hétérogénéité inter-patient est certainement due aux mutations somatiques présentes dans les cellules tumorales. Si l'on compare plusieurs cancers du sein ou du côlon, seulement une petite fraction des mutations somatiques est présente dans les mêmes gènes drivers. Même quand les gènes affectés sont identiques, les types de variations peuvent être différents, ou impacter des domaines distincts de la protéine, conduisant à des conséquences différentes sur les propriétés cellulaires (Cancer Genome Atlas Network, 2012; Stephens et al., 2012; Wood et al., 2007). Cette hétérogénéité inter-patients est un obstacle majeur à l'uniformisation des traitements et démontre la nécessité d'individualiser les traitements en fonction des caractéristiques génétiques des tumeurs de chaque patient (Vogelstein et al., 2013).

I.2.c. Voies de signalisation impliquées dans l'oncogenèse

Un moyen de réduire la complexité du génome des cellules cancéreuses est de "prendre de la hauteur", pour ne plus considérer les gènes individuellement mais au sein de leur voie de signalisation ou de régulation. Il apparaît alors que les gènes porteurs de mutations *drivers* font partie d'un nombre limité de voies. Vogelstein et collaborateurs en recensent 12 pouvant être classées en 3 catégories, selon le processus cellulaire (Figure 3) :

- **La destinée des cellules** : il existe une opposition entre la division et la différenciation cellulaire. Les cellules souches qui se divisent pour peupler les tissus normaux ne se différencient pas et vice-versa. L'homéostasie tissulaire est basée sur un équilibre entre division et différenciation. De nombreuses altérations génétiques des cellules cancéreuses perturbent cet équilibre en favorisant la division. Elles concernent les voies APC/WNT, Hedgehog (HH) et NOTCH.

Lors du développement normal des tissus, l'aiguillage entre division et différenciation n'est pas commandé par des mutations comme dans les cellules cancéreuses, mais par des modifications épigénétiques, affectant le programme transcriptionnel des cellules. Les mutations des voies impliquées dans la modification de la chromatine et la régulation de la transcription peuvent donc également être classées dans cette catégorie.

- **La survie cellulaire** : certaines des mutations acquises par les cellules cancéreuses leur permettent de survivre et proliférer en l'absence de facteurs de croissance ou dans des conditions défavorables, comme en présence de signaux antiprolifératifs ou une concentration insuffisante en nutriments. Cela leur confère un avantage sélectif par rapport aux cellules saines. Il s'agit souvent de mutations activatrices de proto-oncogènes codant pour des récepteurs aux facteurs de croissance eux-mêmes (EGFR, HER2, PDGFR...) ou des intermédiaires qui relaient le signal à l'intérieur des cellules (KRAS, BRAF, PIK3CA...).
Les voies contrôlant le cycle cellulaire et l'apoptose jouent également un rôle important dans la survie cellulaire. Des gènes *drivers* comme *CDKN2A*, *MYC* ou *BCL2*, directement impliqués dans ces processus, sont souvent mutés dans les cellules cancéreuses.
- **La maintenance du génome** : l'exposition des cellules à diverses substances, comme les ROS (*Reactive Oxygen Species*), les rayonnements UV, ou la fumée de cigarette, provoque des altérations de l'ADN. Même en l'absence de ces substances, la réplication de l'ADN ou la division cellulaire sont sources d'erreur (Kunkel, 2009). Il existe des systèmes de contrôle qui permettent de détecter les altérations de l'ADN, ralentir la progression de ces cellules dans le cycle cellulaire et de corriger ces erreurs. Si les altérations ne sont pas réparables, ils déclenchent l'apoptose pour éliminer les cellules endommagées (Ljungman and Lane, 2004; Medema and Macůrek, 2012; Zhou and Elledge, 2000). Les mutations abolissant la fonction des gènes impliqués dans ces voies de contrôle et de réparation de l'ADN autorisent les cellules tumorales à survivre et à se diviser en présence d'altérations qui auraient dû entraîner leur mort par apoptose.

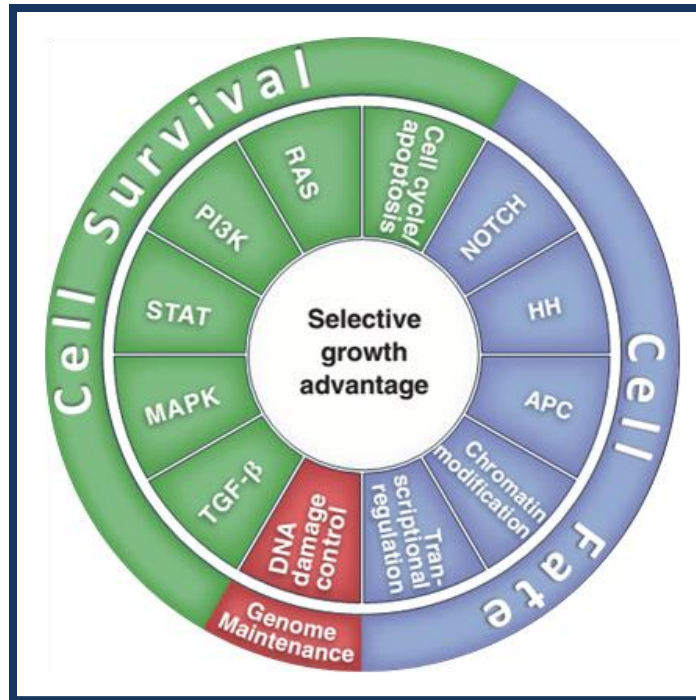


Figure 3: Principales voies de signalisation et processus cellulaires impliqués en cancérologie. (adapté de Vogelstein et al., 2013). HH: Hedgehog.

En considérant les voies physiopathologiques plutôt que les gènes individuellement, et le fait que les mutations des gènes d'une même voie de signalisation sont généralement mutuellement exclusives, c'est à dire qu'elles n'apparaissent pas dans la même tumeur (Ciriello et al., 2012; Yeang et al., 2008), il n'est pas surprenant de constater que différentes mutations dans des gènes d'une même voie, conduisent aux mêmes avantages sélectifs. Cela explique l'hétérogénéité génétique existant entre différentes tumeurs d'un même organe et de même type histologique. En effet, une tumeur pulmonaire peut présenter une mutation d'un récepteur à un facteur de croissance comme *EGFR*, la rendant capable de proliférer en l'absence d'EGF (*Epidermal Growth Factor*). Une seconde tumeur pulmonaire peut présenter une mutation activatrice du gène *KRAS*, qui transmet le signal d'EGFR à l'intérieur de la cellule. Une troisième tumeur pulmonaire peut présenter une mutation inactivatrice du gène *NF1*, dont la fonction est d'inactiver *KRAS*. Les mutations de ces trois gènes favorisent la prolifération cellulaire par la suractivation anormale d'une seule voie de signalisation (RAS/MAPK).

II. Les prédispositions génétiques au cancer

La majorité des cancers sont sporadiques et résultent de l'accumulation de plusieurs mutations somatiques *drivers* au sein d'une même population cellulaire. Individuellement, chaque mutation ne confère qu'un faible avantage prolifératif aux cellules (en moyenne 0,4 % d'augmentation entre le taux de survie et de mort cellulaire). Ce n'est que lorsqu'une même cellule a accumulé plusieurs de ces mutations (entre 5 et 8 événements moléculaires seraient nécessaires) qu'elle obtient un réel avantage sélectif et peut développer une masse tumorale (Vogelstein et al., 2013). Le développement d'un cancer sporadique est donc un processus à long terme et se manifeste la plupart du temps à un âge avancé.

On considère généralement que 5 à 10 % des cancers se développent dans un contexte de prédisposition héréditaire. Près de 400 000 nouveaux cas de cancers étant diagnostiqués chaque année en France, entre 20 000 et 40 000 cas sont imputables à une prédisposition génétique (INCa, 2015a). Elle se transmet le plus souvent selon un mode autosomique dominant avec une pénétrance forte mais incomplète et une expressivité variable. Il existe tout de même des prédispositions génétiques au cancer se transmettant selon un mode autosomique récessif comme la polypose due à des mutations bi-alléliques du gène *MUTYH* (Sampson et al., 2003).

Des formes familiales ont été décrites pour la plupart des localisations mais la proportion est variable selon les types de cancers. Chez l'adulte, la proportion de prédisposition génétique est faible dans certains cancers, comme les cancers du poumon ou ORL. Elle est plus importante dans les cancers du sein, de l'ovaire et du colon (Rahman, 2014). Le syndrome sein-ovaire représente 2 à 5% des cancers du sein et 15 à 20 % des cancers de l'ovaire. Deux à 3 % des cancers colorectaux sont imputables au syndrome de Lynch (INCa, 2015b). Chez l'enfant, les formes héréditaires de cancer se manifestent généralement dans un contexte de prédisposition syndromique. La localisation et le type de tumeur diffèrent des cancers de l'adulte. Les plus fréquentes sont les tumeurs du système nerveux central ou périphérique, les rétinoblastomes, les néphroblastomes et les sarcomes.

Pour qu'il y ait prédisposition au cancer, il suffit qu'une altération génétique *driver* soit présente au niveau germlinal. Le cancer lui-même résultera alors de l'acquisition d'altérations supplémentaires successives dans un ou plusieurs clones cellulaires d'un tissu particulier. La première mutation favorisant la transformation tumorale étant présente dans toutes les cellules de l'organisme, toute mutation somatique « *driver* » supplémentaire dans n'importe quelle cellule viendra ajouter son effet à celui de la mutation constitutionnelle. La présence chez un individu d'une mutation constitutionnelle d'un gène de prédisposition au cancer (GPC) le rend donc susceptible de développer des tumeurs plus fréquemment et plus précocement que dans la population générale.

Sur le plan clinique, certains éléments sont particulièrement évocateurs d'une prédisposition au cancer :

- **La précocité de survenue du cancer:** les tumeurs héréditaires apparaissent avant l'âge moyen de survenue dans la population générale, le plus souvent avant l'âge de 51 ans.
- **L'histoire familiale:** il existe de nombreux sujets atteints de cancer dans la même branche parentale (paternelle ou maternelle). La pénétrance incomplète, à l'origine de "sauts de génération", impose de prendre en compte les antécédents de cancer sur plusieurs générations. La taille de la famille est également à prendre en considération. En effet, une agrégation familiale d'un cancer fréquent en population générale et sans âge jeune ne signifie pas forcément qu'il s'agit d'une forme Mendélienne de cancer.
- **Les tumeurs multiples, multifocales ou bilatérales:** la mutation responsable de la prédisposition étant présente dans toutes les cellules de l'organisme, les sujets concernés sont susceptibles de développer plusieurs tumeurs simultanément ou dans un intervalle de temps réduit. Ils ont également un risque plus élevé de récurrence après un premier cancer.
- **L'agrégation de cancers rares:** la présence de plusieurs cancers rares dans une même famille est suspecte d'une prédisposition génétique. Dans certains cas, la présence d'une seule forme rare peut orienter vers une prédisposition. C'est le cas du cancer du sein chez l'homme, ou la présence d'un homme atteint dans la famille est une indication aux tests de prédisposition génétiques.

Les situations dans lesquels le phénotype est le plus marqué (cancer très précoce, plusieurs sujets atteints dans la même branche familiale, multiplicité des cancers chez un même patient ou récurrence précoce) peuvent être qualifiés de "**phénotype extrême**". La sélection de ces cas permet de se placer dans la situation où une prédisposition génétique est la plus probable. En l'absence de mutation dans les GPC connus, l'approfondissement de l'étude de leurs caractéristiques génétiques par séquençage d'exome pourrait permettre d'identifier de nouveaux GPC.

II.1. Les gènes impliqués dans la prédisposition au cancer

La base de données COSMIC (Catalog Of Somatic Mutation In Cancer) (Futreal et al., 2004) recense 107 gènes drivers dans le développement d'un cancer lorsqu'ils présentent des mutations constitutionnelles. En considérant les gènes dont une mutation entraîne l'apparition d'un cancer chez plus de 5 % des personnes porteuses, ou avec des mutations rares conférant un risque modéré à élevé de développer un cancer (supérieur à 2 fois par rapport à la population

générale), N. Rahman rapporte une liste de 114 GPC (Rahman, 2014). Pour la plupart de ces gènes, le risque et la pénétrance sont beaucoup plus élevés que ces critères minimaux. Plusieurs autres publications dressent des liste de GPC (Garber and Offit, 2005; Vogelstein and Kinzler, 2004; Vogelstein et al., 2013). Elles sont synthétisées dans l'annexe 1.

L'apparition du cancer chez un patient prédisposé est souvent liée à une mutation constitutionnelle hétérozygote d'un gène suppresseur de tumeur, suivie d'un deuxième événement somatique entraînant la perte de l'allèle sain. Cela correspond au « *two-hit model* » proposé par AG. Knudson en 1971 et complété par DE. Comings en 1973 (Figure 4) (Comings, 1973; Knudson, 1971). Ce modèle rend compte de la fréquence élevée des formes bilatérales de rétinoblastome survenant dans les premiers mois de vie dans les cas familiaux en opposition aux formes unilatérales et plus tardives des cas sporadiques. Il a conduit à l'émergence du concept de gène suppresseur de tumeur et à l'identification du premier d'entre eux, le gène *RB1* dans la prédisposition au rétinoblastome.

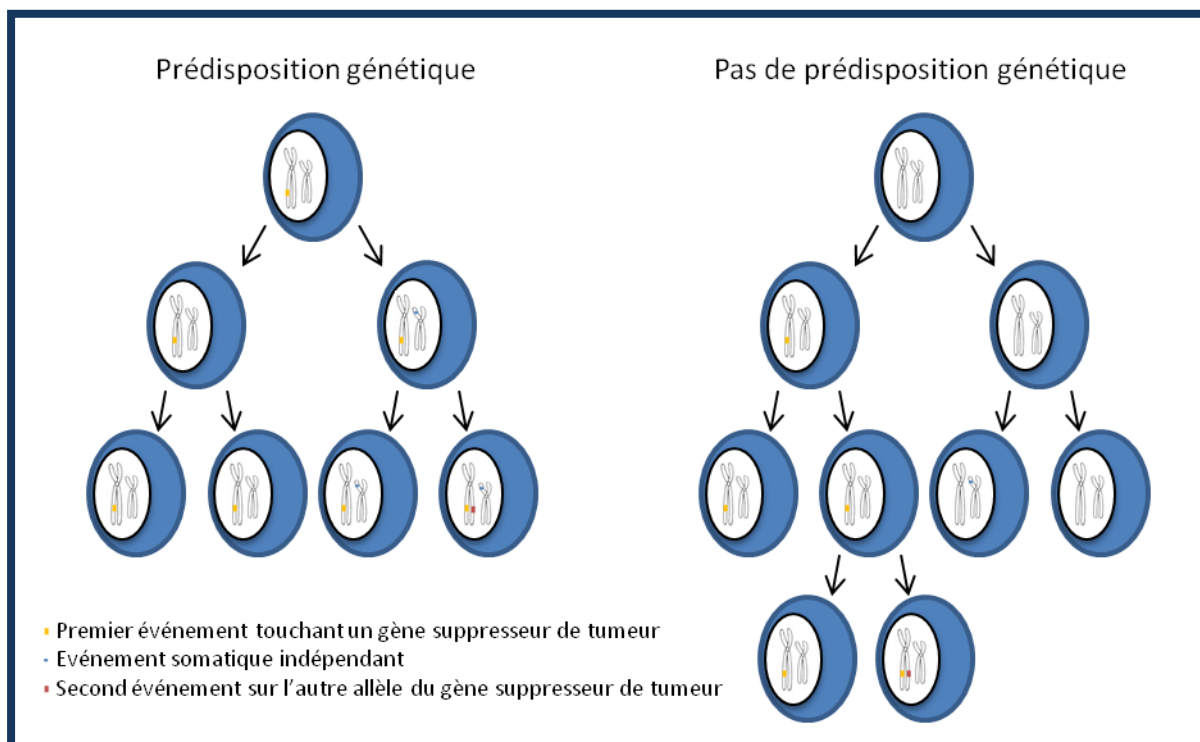


Figure 4:« Two hits model ». Dans le cas d'une prédisposition génétique au cancer impliquant un gène suppresseur de tumeur, la première mutation est présente dans toutes les cellules de l'organisme. Un seul événement supplémentaire sur le second allèle dans n'importe quelles cellules est nécessaire pour favoriser l'apparition du cancer. Cela explique la précocité de la maladie chez les personnes prédisposées.

Depuis, de nombreux autres gènes dont la perte de fonction prédispose au cancer ont été identifiés. Certains sont de véritables gènes suppresseurs de tumeur intervenant dans les processus d'apoptose ou de contrôle du cycle cellulaire (*CDKNA2*, *TP53*) ou inhibant des voies de prolifération cellulaire (*APC*, *PTEN*). Beaucoup de GPC sont des gènes de réparation de l'ADN impliqués dans la réparation des cassures double brin (*ATM*, *BRCA1/2*), le système MMR (*MLH1*,

MSH2, *MSH6*, *PMS2*, *EPCAM*), ou le NHEJ (*DCLRE1C*). Une inactivation bi-allélique de ces gènes est souvent nécessaire pour que le phénotype s'exprime en suivant le modèle de Knudson et Comings. Cependant, la tolérance biologique aux variations de l'expression des gènes suppresseurs de tumeur est variable selon la protéine concernée, pouvant donner lieu à des modes d'action ne nécessitant pas la perte complète de l'expression du gène. Le contexte cellulaire et génétique peut également influencer l'effet de la mutation d'un gène.

Dans certains cas, la perte d'un seul allèle d'un gène suppresseur de tumeur a un pouvoir cancérigène plus important que la perte totale de son expression. Il existe également des situations où une faible diminution de l'expression d'un gène est suffisante pour favoriser la cancérogénèse. Ces phénomènes sont qualifiés d'haploinsuffisance obligatoire et de quasi-insuffisance (Berger et al., 2011). L'haploinsuffisance obligatoire peut avoir lieu si l'inactivation bi-allélique du gène en question conduit à la mort ou la sénescence de la cellule. La pression de sélection qui s'exerce lors de la tumorigénèse va alors favoriser les cellules n'ayant perdu qu'un seul allèle du gène. Une autre possibilité est que la mutation d'un seul allèle ait un effet dominant négatif. Dans ce cas, la protéine mutante interfère avec la protéine normale produite à partir du second allèle et perturbe son activité. La fonction normale du gène suppresseur de tumeur étant déjà totalement perdue avec la mutation d'un seul allèle, il n'y a pas de pression de sélection en faveur de la perte du second allèle (Berger et al., 2011). Une troisième possibilité est qu'un gène suppresseur de tumeur soit haploinsuffisant et que la perte d'une seule copie soit suffisante pour favoriser la transformation tumorale. Il a également été démontré sur un modèle murin qu'une faible diminution de l'expression du gène suppresseur de tumeur *PTEN* peut avoir un pouvoir cancérigène, introduisant ainsi le concept de quasi-insuffisance (Alimonti et al., 2010). Cela montre que, bien que le modèle de Knudson et Comings ait permis la caractérisation de nombreux gènes suppresseurs de tumeur, il ne doit pas être érigé en dogme et d'autres mécanismes sont à prendre en considération pour l'identification de gènes suppresseurs de tumeur.

Il existe des gènes responsables d'une prédisposition au cancer à la fois chez les porteurs de mutations mono et bi-alléliques, avec des manifestations plus sévères en condition récessive. Par exemple, les porteurs de mutations bi-alléliques de *BRCA2*, *PALB2*, *MLH1*, *MSH2*, *MSH6* et *PMS2* ont un risque élevé de tumeurs pédiatriques alors que les porteurs de mutations mono-alléliques ont un risque élevé de cancer à l'âge adulte (Rahman and Scott, 2007).

Certains gènes sont plus difficiles à classer en gènes suppresseurs de tumeur ou de réparation. Il semble que la prédisposition au cancer soit une complication des manifestations des syndromes auxquels ils sont associés. Par exemple, la déficience du gène *FAH* dans la tyrosinémie de type I est responsable de l'accumulation dans les cellules d'un métabolite mutagène, le fumarylacétoacétate, qui favorise le développement de carcinomes hépatocellulaires (Jorquera and Tanguay, 2001). Un autre exemple est l'inactivation bi-allélique du gène *DOCK8*, responsable du syndrome hyper-IgE autosomique récessif (MIM#243700). L'immunodéficience qui le caractérise facilite l'infection par des virus cutanés et entraîne une susceptibilité accrue au développement de carcinomes épidermoïdes (Zhang et al., 2009).

Les proto-oncogènes responsables de prédisposition au cancer sont plus rares. La plupart d'entre eux codent pour des kinases rendues constitutionnellement actives par la mutation (Rahman, 2014). Comme pour les gènes suppresseurs de tumeurs, dans certaines situations, une trop forte activité d'un oncogène peut induire la sénescence ou l'apoptose au lieu de favoriser la prolifération cellulaire (Evan et al., 1992; Serrano et al., 1997).

II.2. Les principales prédispositions génétiques au cancer

Les prédispositions génétiques les plus fréquentes sont le syndrome sein-ovaire (MIM# 604370, MIM#612555), la polypose adénomateuse familiale (PAF) (MIM## 175100) et le syndrome de Lynch ou HNPCC (Hereditary Non-Polyposis Colorectal Cancer) (MIM# 120435).

Des altérations constitutionnelles hétérozygotes des gènes *BRCA1* et *BRCA2*, impliqués dans la réparation des lésions double-brin de l'ADN, sont la principale cause du syndrome sein-ovaire. Elles sont associées à un risque très élevé de cancer du sein et/ou de l'ovaire avec un risque cumulé de cancer du sein à l'âge de 80 ans de 72 % et 69 % et de cancer de l'ovaire de 44 % et 17 % pour les femmes porteuses de mutations de *BRCA1* ou *BRCA2* respectivement (Kuchenbaecker et al., 2017). Les femmes porteuses de ces altérations ont également un risque augmenté de développer un second cancer sur le sein controlatéral. Cependant, les mutations des gènes *BRCA1* et *BRCA2* ne permettent pas d'expliquer plus de la moitié des syndromes sein-ovaire. D'autres gènes ont été identifiés en lien avec ce syndrome. En 2017, le Groupe Génétique et Cancer (GGC) a établi un panel de 13 gènes reconnus d'utilité clinique lors de la suspicion d'un syndrome sein-ovaire (*BRCA1*, *BRCA2*, *PALB2*, *TP53*, *CDH1*, *PTEN*, *RAD51C*, *RAD51D*, *MLH1*, *MSH2*, *MSH6*, *PMS2* et *EPCAM*) (GGC-Unicancer, 2017). Le cancer du sein présente une incidence élevée dans la population générale. Il peut être compliqué de distinguer les formes héréditaires familiales des concentrations familiales fortuites. Les éléments à prendre en compte sont le nombre de cas de cancers du sein chez des apparentés de premier ou de deuxième degré dans la même branche parentale, la précocité de survenue du cancer du sein (40 ans ou moins), l'atteinte bilatérale, la présence de cancer(s) de l'ovaire et l'observation éventuelle de cancer du sein chez l'homme dans la famille.

La PAF est une maladie à transmission autosomique dominante liée à des mutations du gène *APC* (Nishisho et al., 1991). Elle se manifeste classiquement par l'apparition de centaines de polypes dans le côlon dès l'adolescence. Individuellement, ces polypes ne sont pas plus susceptibles de se transformer en cancer que chez une personne non atteinte de PAF mais le nombre élevé augmente le risque que l'un d'eux devienne cancéreux. Si on ne prend pas de mesure préventive (résection prophylactique du côlon), le risque de cancer est presque de 100% avant 40 ans dans les formes classiques (Half et al., 2009). Néanmoins, des formes plus modérées existent, avec des risques de cancer plus faibles. Il existe également une autre forme de PAF, moins sévère, liée à des mutations du gène *MYH*. Cette forme, autosomique récessive, se manifeste par un début plus tardif, un nombre plus faible de polypes, et un risque de cancer moins important.

Le syndrome de Lynch est dû à des mutations des gènes du système de réparation des mésappariements des bases lors de la réplication de l'ADN (système MMR: MisMatchRepair, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM*). Il engendre la formation de polypes dans la paroi du côlon mais pas en aussi grand nombre que dans le cas de la PAF. Le risque de développer un cancer colorectal est de l'ordre de 10 % à 50 ans et 40 % à 70 ans (ERISCAM, 2010). Il représente 2 à 3% de l'ensemble des cancers colorectaux. Les femmes présentent aussi un risque accru de cancer de l'endomètre et les personnes atteintes peuvent développer d'autres cancers (voies biliaires, voies urinaires, ovaire, intestin grêle, estomac). Cette prédisposition génétique est discutée lorsqu'on diagnostique un cancer colorectal chez une personne de moins de 50 ans, chez un patient de moins de 60 ans dont la tumeur présente une instabilité des microsatellites ou lorsque des apparentés ont présenté un cancer appartenant au spectre des cancer de ce syndrome (INCa, 2015b).

II.3. Recherche des bases moléculaires des prédispositions génétiques au cancer

Depuis 2010, la mise en place de techniques de séquençage haut débit, communément dénommées NGS (Next Generation Sequencing), dans les laboratoires d'oncogénétique a permis de réduire notablement les coûts et les délais de séquençage. Cela a rendu possible l'analyse simultanée de groupes de gènes (ou panels) connus pour être impliqués dans les formes mendéliennes de cancer. Cette technique est devenue la pratique courante pour identifier les bases moléculaires d'une prédisposition au cancer. Les différents panels de gènes peuvent être différents selon les laboratoires, malgré un effort d'homogénéisation récent, et ont un rendement diagnostique limité. Une publication récente portant sur plus de 10 000 patients montre que le séquençage de panels de gènes selon les indications cliniques permet de poser un diagnostic moléculaire dans seulement 9% des cas en moyenne (Susswein et al., 2016). Avec 18,7 % de cas résolus, le panel dédié aux cancers du spectre du syndrome de Lynch a le meilleur rendement diagnostique, certainement en raison de critères cliniques bien définis et de l'étude préalable des caractéristiques tumorales (instabilité des microsatellites et/ou expression des protéines MMR par immunohistochimie). Ces chiffres concordent avec ceux rapportés par l'Institut National du Cancer (INCa) pour l'année 2016 avec une altération génétique liée au syndrome sein-ovaire retrouvée chez seulement 10% des cas index testés, et dans 19% des cas pour le syndrome de Lynch (INCa, 2016). Cela montre que malgré l'apport du NGS, l'analyse des GPC connus ne permet d'expliquer qu'une petite proportion des situations où une prédisposition génétique est suspectée.

II.4. Conséquences de la découverte d'une variation pathogène dans un gène de prédisposition au cancer

Les arguments généalogiques et individuels permettent de retenir, avec plus ou moins de certitude, une prédisposition génétique, voire d'estimer le risque de cancer et ainsi de proposer aux apparentés d'un sujet atteint une prise en charge spécifique avec la mise en place de

mesures de prévention et de dépistage précoce des tumeurs. L'identification de la mutation responsable chez le cas index permet de confirmer le diagnostic de prédisposition et de proposer une analyse pré-symptomatique chez les apparentés. Dans ce cas, l'analyse est simple et rapide puisqu'elle est basée sur la recherche ciblée de l'altération mise en évidence chez le cas index. Un résultat négatif permet de rassurer le patient, son risque de développer un cancer rejoignant celui de la population générale, et lui épargne des mesures de dépistage anxiogènes et coûteuses pour l'assurance maladie.

Au contraire, l'identification des apparentés porteurs du risque génétique permettra de préciser le risque de cancer et de proposer une prise en charge personnalisée visant à détecter précocement l'apparition d'une tumeur. Elle permet aussi de proposer des mesures de prévention (chirurgie prophylactique) aux individus à très haut risque.

Dans certaines situations, l'identification de la mutation responsable de la prédisposition peut également avoir un impact sur le traitement. Par exemple, les chimiothérapies à base de sels de platine ont démontré leur utilité dans le traitement des cancer du sein (Byrski et al., 2012; Tutt et al., 2018) et de l'ovaire (Walsh, 2015) liés à une mutation des gènes *BRCA1* et *BRCA2*. De nombreuses thérapies ciblées ont été développées ces dernières années. Leur principe est d'exploiter la connaissance de la cause sous-jacente du cancer pour identifier les vulnérabilités spécifiques à la tumeur pouvant être la cible d'une thérapeutique. Le modèle le plus simple concerne les cancers liés à des mutations gain de fonction de proto-oncogènes, pouvant être directement inhibés par des molécules tels que l'Imatinib (*KIT* et *PDGFRA*), le Vandétanib (*RET*) et le Foretinib (*MET*) (Bachet et al., 2013; Logan, 2013; Wells et al., 2010). Il est plus compliqué de rétablir la fonction d'un gène suppresseur de tumeur. En revanche, si la mutation conduit à l'activation anormale d'une voie métabolique, il est possible de cibler d'autres membres de cette voie afin de l'inhiber. C'est le cas de l'Everolimus, un inhibiteur de MTOR utilisé pour traiter les astrocytomes dans la sclérose tubéreuse de Bourneville due à des mutations des gènes *TSC1* et *TSC2* (Józwiak et al., 2012). Un autre exemple est l'utilisation d'un inhibiteur de la voie Hedgehog (Vismodegib) pour traiter les carcinomes basocellulaires des patients atteints du syndrome de Gorlin dû à la mutation du gène *PTCH1* (Tang et al., 2012).

Une des approches thérapeutiques les plus innovantes est l'induction de la létalité synthétique. Deux gènes ont une relation de létalité synthétique quand leur inhibition simultanée cause la mort cellulaire. La perte de l'un ou l'autre des gènes seul n'entraîne pas la mort cellulaire et peut au contraire constituer un avantage de survie. Cette approche est très prometteuse chez les patients prédisposés au cancer lié aux gènes *BRCA1* ou *BRCA2*: l'enzyme PARP (Poly ADP-Ribose Polymerase) est impliquée dans le processus de réparation de l'ADN par excision de base. Son inhibition entraîne la formation de lésions, normalement réparées par recombinaison homologue, processus impliquant *BRCA1* et *BRCA2*. Les tumeurs de ces patients ayant généralement une perte du second allèle de *BRCA1* ou *BRCA2*, l'inhibition de la voie PARP va entraîner la mort des cellules tumorales mais pas des cellules saines, qui ne sont porteuses que de mutations monoalléliques (Farmer et al., 2005; Fong et al., 2009). Le développement de cette approche thérapeutique se poursuit pour d'autres GPC (Brough et al., 2011). Plusieurs études ont montré une amélioration de la survie sans cancer grâce aux inhibiteurs de PARP,

principalement dans les cancers de l'ovaire. L'Olaparib a d'ailleurs une autorisation de mise sur le marché en monothérapie pour le traitement d'entretien des patientes adultes atteintes d'un cancer épithélial séreux de haut grade de l'ovaire, récidivant et sensible au platine avec une mutation du gène *BRCA* (germinale et/ou somatique) et qui sont en réponse (complète ou partielle) à une chimiothérapie à base de platine (INCa, 2017). Ces molécules sont aussi à l'étude pour le traitement des cancers du sein et de la prostate.

III. Le séquençage haut débit

Le séquençage par la méthode de Sanger automatisée est considéré comme la technologie de séquençage de première génération. Les méthodes de deuxième génération, qualifiées de séquençage haut débit ou souvent désignées par l'acronyme anglais NGS (Next Generation Sequencing) sont une révolution technologique sans précédent dans le domaine de la génétique que ce soit en recherche fondamentale ou appliquée ou en routine diagnostic. Les évolutions spectaculaires de la technologie en termes de capacité de séquençage et la diminution des coûts associés permettent aujourd'hui de séquencer un génome humain complet en quelques jours pour un coût moyen de 1500 euros, alors que le premier séquençage du génome humain selon la méthode de Sanger nécessita plus de 10 ans de travail et presque 3 milliards de dollars. Cette énorme capacité de séquençage a élargi les champs d'application, au-delà de la simple détermination de l'ordre des bases. Par exemple, les études d'expression génique sur puce sont maintenant remplacées par des méthodes basées sur le NGS capables dans la même analyse d'identifier et de quantifier des transcrits rares, de donner des informations sur l'épissage alternatif et de détecter des variations candidates. Le ChIP-Seq (Chromatin Immuno Precipitation - Sequencing) permet de cartographier tous les sites de liaison à l'ADN d'une protéine à l'échelle du génome. La capacité à séquencer le génome de nombreux organismes apparentés a permis de conduire des études comparatives et évolutionnistes à grande échelle, inimaginables sans cette technologie.

L'application la plus large du NGS est sans doute le re-séquençage du génome humain afin de comprendre comment les variations génétiques identifiées affectent notre santé et influent sur le développement de maladies (Koboldt et al., 2013; Metzker, 2010). Le NGS a permis le transfert d'une médecine génétique basée sur l'analyse ciblée de quelques gènes vers des approches pan-génomiques, comme le séquençage d'exome (analyse de toutes les séquences codantes) ou du génome, en passant par le transcriptome (séquençage de tous les transcrits) ou le méthylome (étude de la méthylation du génome). Cette révolution technique du NGS a nécessité d'importants développements en bioinformatique pour le traitement des données brutes générées par le séquençage. Aujourd'hui, le véritable défi de la génétique médicale n'est plus la détection des mutations, mais l'évaluation de leur impact biologique et leur implication dans la pathologie étudiée, pour mieux orienter le traitement à l'avenir.

III.1. Aspects techniques du séquençage haut débit

L'énorme augmentation de débit du NGS par rapport au séquençage de première génération repose sur deux principes majeurs : l'amplification clonale et le séquençage simultané de plusieurs millions de petits fragments d'ADN. La séquence étudiée est ensuite reconstituée *in silico* en alignant les lectures par rapport à un génome de référence, s'il existe, ou par assemblage des séquences chevauchantes.

III.1.a. Préparation des librairies

La première étape de l'analyse consiste à produire des millions de courts fragments d'ADN, copies des régions d'intérêt que l'on veut étudier. Cette banque d'ADN est appelée librairie (anglicisme dérivé de "library"). L'ADN peut être fragmenté par sonication ou par des enzymes de restriction. Si l'on veut séquencer la totalité du génome, seule la fragmentation de l'ADN génomique sera nécessaire. Mais on ne s'intéresse souvent qu'à une petite partie du génome. Par exemple, uniquement la partie codante (exome), ou seulement certains gènes (panel de gènes). Il sera alors nécessaire de réaliser une étape de sélection (ou enrichissement) des régions d'intérêt.

Il existe trois méthodes d'enrichissement:

- **Enrichissement par PCR:** cette méthode consiste à amplifier les régions d'intérêt par une ou plusieurs réactions PCR multiplex, c'est à dire l'amplification simultanée de nombreux fragments différents dans la même réaction. Une étape préliminaire de fabrication des amorces est nécessaire. Il existe également des techniques de PCR en émulsion faisant appel à la micro-fluidique. Dans ce cas, chaque réaction de PCR est isolée dans une microgouttelette et toutes les amplifications ont lieu simultanément dans le même tube.
- **Enrichissement par capture d'hybrides:** l'ADN est d'abord fragmenté par sonication ou digestion enzymatique, puis les régions d'intérêt sont capturées grâce à des sondes complémentaires biotinylées. Les hybrides sont ensuite purifiés grâce à des billes de streptavidine.
- **Enrichissement par circularisation:** l'ADN génomique est fragmenté puis enrichi grâce à des sondes constituées d'une séquence universelle flanquée de séquences spécifiques de la région d'intérêt. Les fragments d'ADN capturés sont ensuite amplifiés grâce à la séquence universelle des sondes.

Au cours de ces étapes d'enrichissement, on ajoute à chaque extrémité des fragments d'ADN de la librairie, des séquences adaptatrices qui vont permettre leur fixation sur un support (bille ou lame de verre) en vue de leur amplification clonale et du séquençage. Les séquences adaptatrices sont spécifiques du séquenceur qui sera utilisé. On peut aussi associer à chaque fragment de la librairie une courte séquence d'ADN synthétique identifiant l'échantillon. L'utilisation de ces séquences, appelées index, permet de séquencer plusieurs échantillons au cours d'une même analyse.

III.1.b. Amplification clonale

Avant le séquençage des librairies, il est nécessaire de multiplier le nombre de copies de chaque fragment d'ADN séparément. Ce processus appelé amplification clonale peut se faire par

PCR en émulsion. Chaque fragment d'ADN à séquencer est fixé sur une microbille grâce aux séquences adaptatrices. Puis, l'émulsion permet d'isoler chaque bille dans un microréacteur et une PCR amplifie le fragment d'ADN pour recouvrir la bille d'un millier de copies du fragment d'ADN initial. Les billes sont ensuite injectées dans les micro-puits d'une puce, dimensionnés pour ne contenir qu'une seule bille (Figure 5A). Chaque puits de la puce sera le siège d'une réaction de séquençage. Cette technique est utilisée par les plateformes de séquençage commercialisées par les sociétés Roche ou Life Technologies.

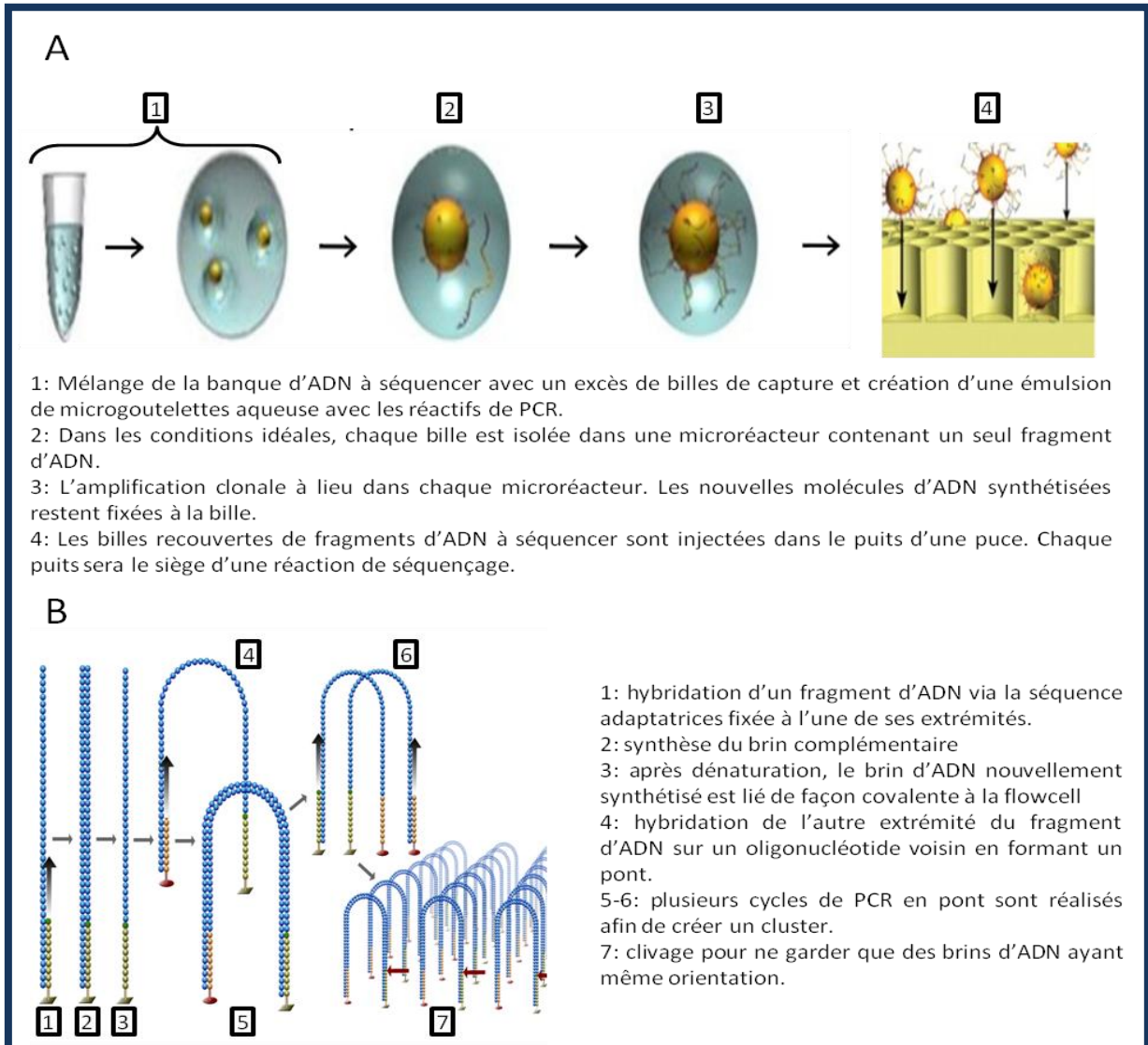


Figure 5: Amplification clonales. A: PCR en émulsion. B: PCR en pont sur support solide

La seconde technique d'amplification clonale consiste à créer des groupes ou "clusters" de fragments d'ADN sur un support solide. La banque de fragments d'ADN ainsi constituée est injectée à l'intérieur d'une cellule de verre, appelée *flowcell*, dont la surface est tapissée d'oligonucléotides complémentaires des séquences adaptatrices. Plusieurs cycles de PCR vont permettre de créer environ 1000 copies de chaque fragment d'ADN original lié à la flowcell. Cette méthode d'amplification clonale a été nommée PCR en pont car les brins d'ADN se

recourbent pour s'hybrider aux oligonucléotides voisins liés à la surface de la flowcell (Figure 5B) (Bentley et al., 2008; Fedurco et al., 2006). L'amplification clonale sur support solide est utilisée pour tous les séquenceurs de la société Illumina.

III.1.c. Les différentes plateformes de séquençage haut débit

Dès la fin des années 1990, plusieurs types de séquenceur de deuxième génération ont été développés. La première technologie de NGS fut basée sur le principe du pyroséquençage (Ronaghi et al., 1996, 1998). Dans cette méthode, deux enzymes sont utilisées pour convertir le pyrophosphate libéré lors de l'incorporation d'un nucléotide à l'extrémité d'une chaîne nucléotidique en un signal lumineux (Figure 6A). La séquence est déterminée en mesurant l'intensité de la luminosité après l'injection séquentielle de chaque nucléotide dans le milieu réactionnel. La principale difficulté de cette technique est de déterminer combien de fois le même nucléotide est répété à une position donnée. En effet, bien que l'intensité du signal lumineux soit proportionnelle au nombre de nucléotides incorporés, la relation n'est plus linéaire au delà de 4 ou 5 répétitions. La société 454 Life Science acquise par Roche en 2007 a commercialisé les premiers séquenceurs de deuxième génération basés sur cette technologie.

Suite au succès de 454 Life Science, un certain nombre de technologies de séquençage de deuxième génération ont vu le jour. La plus importante d'entre elle étant certainement la méthode développée par la société Solexa (Voelkerding et al., 2009), plus tard acquise par Illumina. Le séquençage s'effectue par synthèse en utilisant des dNTP terminateurs de chaîne réversible. La position 3'OH de ces nucléotides est occupée par un fluorophore qui doit être clivé avant de pouvoir continuer la synthèse de la chaîne nucléotidique (Voelkerding et al., 2009). Chaque dNTP étant couplé à un fluorophore différent, ils sont injectés simultanément dans la flowcell puis "lavés" avant lecture de la fluorescence de chaque cluster. Plusieurs cycles consécutifs d'injection - lavage - lecture sont réalisés pour déterminer l'enchaînement des bases (Figure 6B). Les premiers séquenceurs Illumina, baptisés Genome Analyzer ne produisaient que de très courtes séquences (jusqu'à 35 bases) mais avaient l'avantage de lire la séquence des deux extrémités des brins d'ADN. On appelle cela le séquençage "paired-end". Après que la séquence d'une des extrémités soit lue, les molécules d'ADN sont retournées par un unique cycle de PCR en pont et une seconde lecture est réalisée à partir de l'autre extrémité. Etant donné que la taille approximative des fragments d'ADN séquencés est connue, l'obtention de données pairées permet d'améliorer la précision de l'alignement des séquences sur le génome de référence, spécialement dans les régions répétées, et facilite la détection de réarrangements, de gènes de fusion ou d'épissages alternatifs.

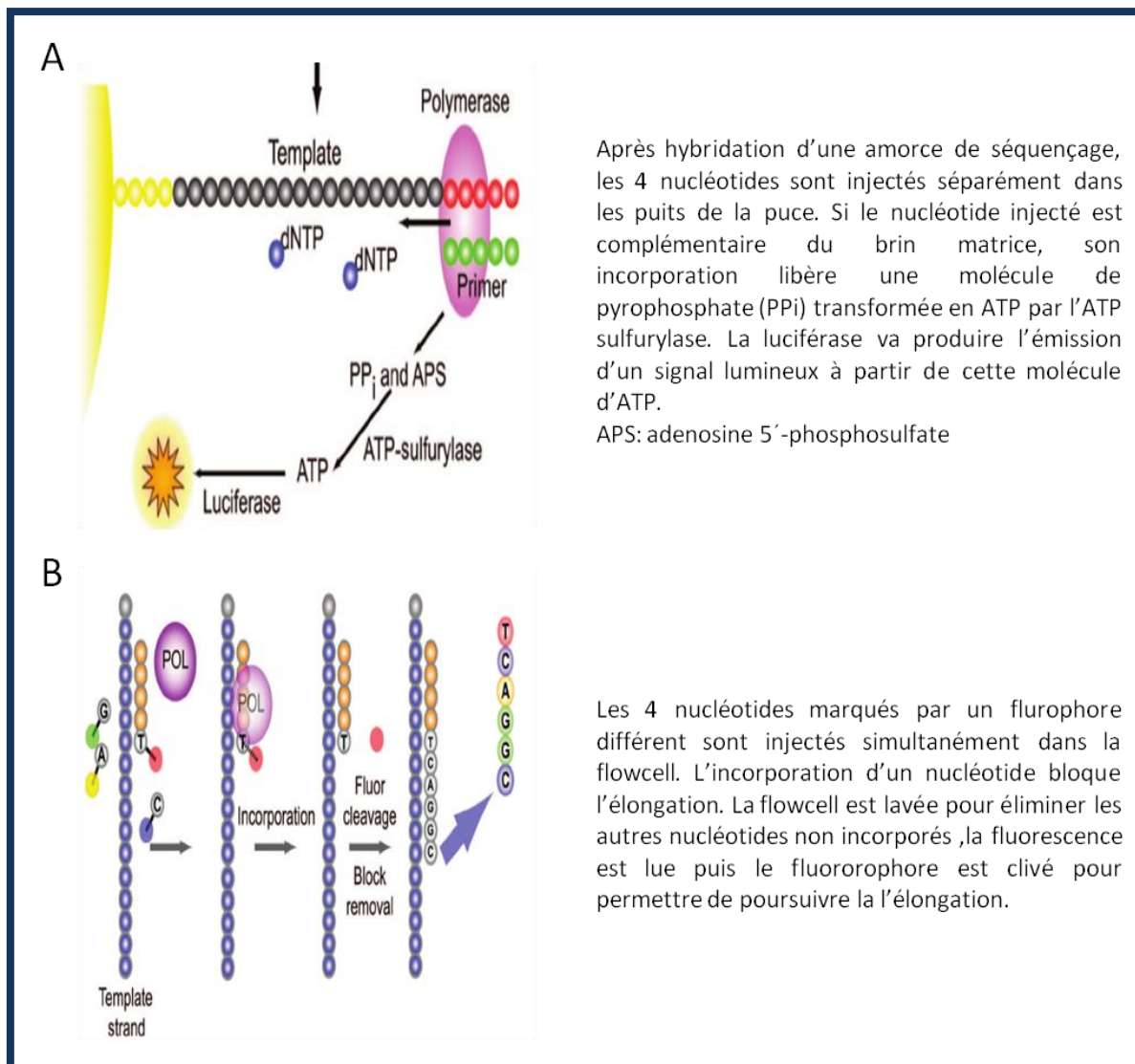
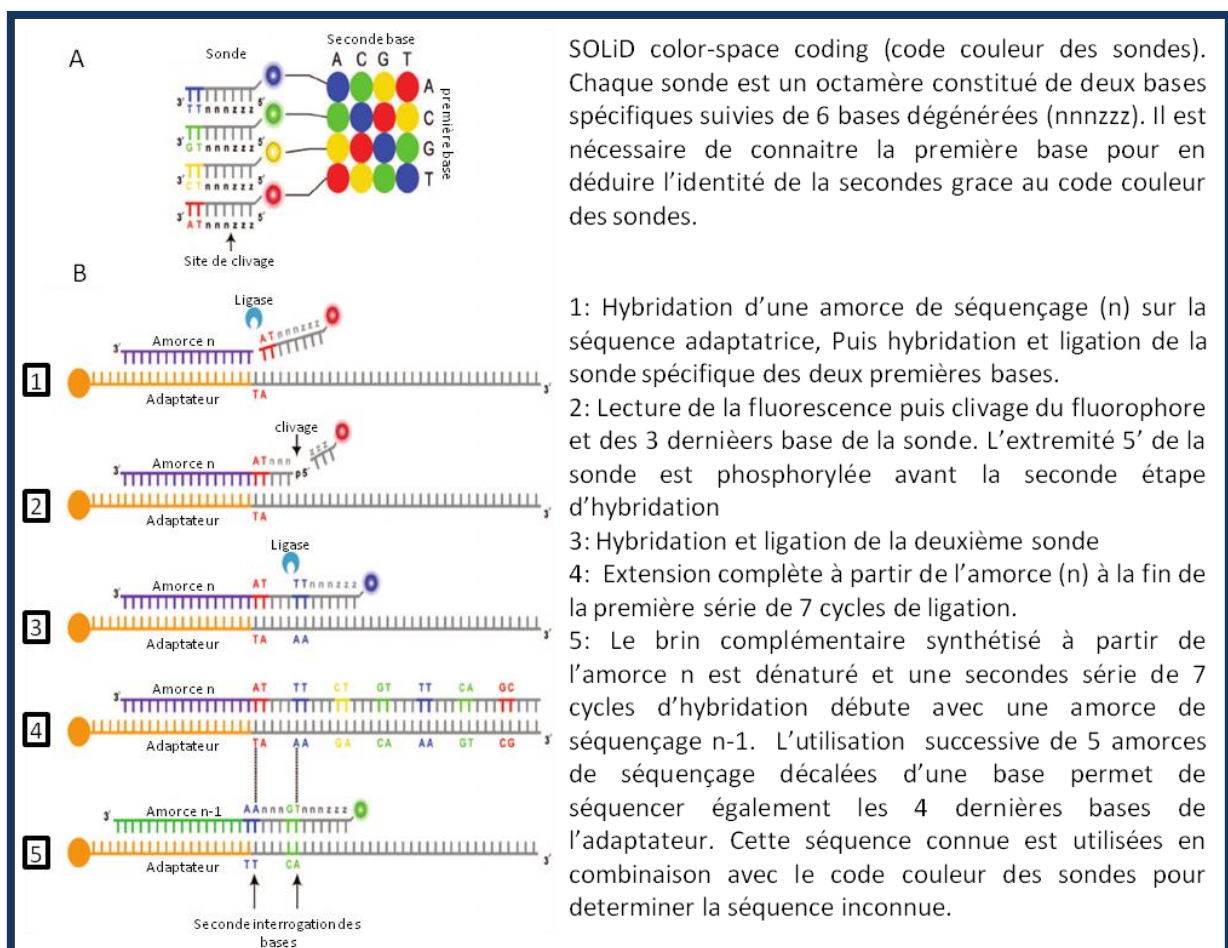


Figure 6: Séquençage par synthèse. A: Méthode par pyroséquençage 454. B: Méthode Illumina par bloqueurs de chaîne réversibles.

La troisième technologie de séquençage par synthèse est celle développée par la société Ion Torrent acquise par Life technologies en 2010. D'une manière analogue au séquençage 454, des billes tapissées d'une population clonale de fragments d'ADN (générées par PCR en émulsion) sont déposées dans les micro-puits d'une puce. Le séquençage par synthèse est réalisé par l'injection séquentielle de chacun des 4 dNTP. L'originalité de cette technique réside dans la détection du nucléotide incorporé. Elle est réalisée par la mesure d'une variation de pH due à la libération de protons H⁺ lors de la polymérisation. Cette méthode permet un séquençage très rapide mais comme la technologie 454, elle peine à déterminer le nombre de nucléotides incorporés dans les séquences d'homopolymères (Glenn, 2011; Rothberg et al., 2011).

Une autre méthode a été développée parallèlement au séquençage par synthèse. Il s'agit du séquençage par ligation. Dans cette stratégie, des sondes de huit bases sont hybridées sur le fragment d'ADN à séquencer puis liées entre elles par une ligase. Chaque sonde est constituée d'un dimère correspondant aux bases à séquencer, suivi de trois bases dégénérées et de trois bases universelles. Les bases dégénérées et universelles sont capables de s'apparier avec les

quatre nucléotides (Figure 7A). Seize sondes correspondant aux seize dimères de bases possibles sont utilisées. Elles sont marquées par quatre fluorochromes différents. Cinq séries de sept cycles sont nécessaires pour déterminer la séquence du brin à séquencer. Chaque cycle est constitué d'une étape d'hybridation et de ligation de la sonde spécifique puis d'une étape de détection de la fluorescence de la sonde avant digestion des trois bases universelles. La digestion de ces bases permet de régénérer l'extrémité 5' nécessaire pour le cycle suivant. A la fin de ces cycles, le brin nouvellement synthétisé est dénaturé. Une nouvelle série de sept cycles d'hybridation est répétée avec un décalage d'une base. A chaque série, la fluorescence des sept dimères est détectée et la séquence de la molécule d'ADN est déterminée par l'interprétation de la combinaison des différents dimères des cinq séries (Figure7B). Cette technologie a l'avantage d'interroger deux fois chaque base, réduisant ainsi le taux d'erreurs de lecture. En 2005, Shendure et ses collaborateurs ont décrit une technologie associant cette méthode de ligation à l'amplification clonale par PCR en émulsion. Elle fut à l'origine de la technologie SOLiD (*Supported Oligonucleotide Ligation and detection*) commercialisée par la société Life technologies (Shendure et al., 2005).



SOLiD color-space coding (code couleur des sondes). Chaque sonde est un octamère constitué de deux bases spécifiques suivies de 6 bases dégénérées (nnzzz). Il est nécessaire de connaître la première base pour en déduire l'identité de la seconde grâce au code couleur des sondes.

- 1: Hybridation d'une amorce de séquençage (n) sur la séquence adaptatrice, Puis hybridation et ligation de la sonde spécifique des deux premières bases.
- 2: Lecture de la fluorescence puis clivage du fluorophore et des 3 dernières base de la sonde. L'extrémité 5' de la sonde est phosphorylée avant la seconde étape d'hybridation
- 3: Hybridation et ligation de la deuxième sonde
- 4: Extension complète à partir de l'amorce (n) à la fin de la première série de 7 cycles de ligation.
- 5: Le brin complémentaire synthétisé à partir de l'amorce n est dénaturé et une seconde série de 7 cycles d'hybridation débute avec une amorce de séquençage n-1. L'utilisation successive de 5 amorces de séquençage décalées d'une base permet de séquencer également les 4 dernières bases de l'adaptateur. Cette séquence connue est utilisée en combinaison avec le code couleur des sondes pour déterminer la séquence inconnue.

Figure 7: Séquençage par ligation selon la technologie SOLiD. A: Les sondes et leur code couleur. B: Déroulement de séquençage

La révolution génomique engendrée par ces technologies de deuxième génération a grandement facilité le séquençage de l'ADN et diminué drastiquement son coût. Les capacités

des séquenceurs ont évolué à un rythme plus rapide que celui décrit par la loi de Moore lors de la révolution informatique: la complexité des microprocesseurs doublait à peu près tous les deux ans, alors que la capacité des séquenceurs a doublé tous les cinq mois entre 2004 et 2010 (Stein, 2010). Plusieurs séquenceurs avec des caractéristiques et des capacités diverses co-existent sur le marché fournissant divers outils aux chercheurs pour concevoir leurs analyses. Toutefois, ces dernières années, le succès des instruments de la société Illumina la place comme leader sur le marché (Greenleaf and Sidow, 2014). Après l'abandon de la commercialisation de la technologie 454 par Roche en 2013, la société Thermo Fisher Scientific, qui commercialise des séquenceurs basés sur la technologie Ion Torrent, est son seul concurrent.

Une troisième génération de séquenceurs basés sur l'analyse de molécules d'ADN unique commence à faire son apparition. Ils sont capables de séquencer de très longues molécules d'ADN en un temps très court sans besoin d'amplification clonale. Cependant, leur taux d'erreur est trop élevé pour être compatible avec la détection de petites variations génétiques. Leur principal intérêt est la création d'assemblages *de novo* ou leur utilisation en combinaison avec des données de deuxième génération pour améliorer leur alignement et la détection de grandes variations structurales (Heather and Chain, 2016).

III.1.d. Traitement bioinformatique des données brutes

Plusieurs étapes d'analyse bioinformatique sont nécessaires pour convertir les données brutes issues des séquenceurs en une liste de variations interprétables. Le processus de traitement des données brutes générées par un séquenceur Illumina appliqué à la détection de variations génétiques peut être divisé en six grandes étapes (Figure 8).

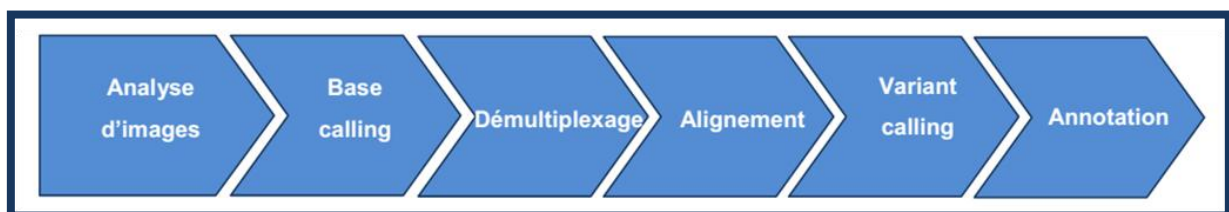


Figure 8: Différentes étapes de l'analyse bioinformatique pour l'identification de variations génétiques.

La première étape consiste à analyser les images réalisées lors du séquençage pour déterminer la position des clusters, leur intensité et leur bruit de fond. La seconde étape, appelée "*base calling*", permet de déterminer chaque base individuellement et de générer des fichiers de séquences brutes au format ".bcl". Ces deux premières étapes sont réalisées en même temps que le séquençage. Elles intègrent la détermination de plusieurs facteurs de qualité des clusters. Les clusters dont la qualité est insuffisante sont exclus lors des premiers cycles de séquençage. Un score de qualité est également calculé pour chaque base. Il s'agit du score de Phred (ou Q score). Il représente, sur une échelle logarithmique, la probabilité que la base appelée soit fautive. Par exemple, un Q score de 30 (Q30) correspond à une probabilité d'erreur de 1 sur 1000, donc à une précision de 99,9% (Tableau 1).

Tableau 1: Relation entre le score de Pred et la probabilité d'erreur de l'appel d'une base.

Score de Phred	Probabilité d'un mauvais appel de base	Précision de l'identification de la base
Q10	1 pour 10	90%
Q20	1 pour 100	99%
Q30	1 pour 1000	99,9%
Q40	1 pour 10000	99,99%

L'étape de démultiplexage permet l'identification des séquences d'index spécifique à chaque échantillon et génère un ou plusieurs fichiers au format ".fastq". Il s'agit de fichiers texte contenant à la fois la séquence nucléique lue pour chaque cluster et le score de Pred associé à chaque base individuellement.

Chaque lecture doit ensuite être alignée sur le génome de référence. Compte tenu du très grand nombre de lectures à aligner, cette étape nécessite des outils informatiques très performants. Il existe plusieurs logiciels d'alignements. La plupart utilisent des méthodes d'indexation du génome de référence pour définir rapidement les positions potentielles des lectures. Ils génèrent des fichiers binaires au format ".bam" (Binary Alignment Map).

L'étape suivante consiste à détecter les variations entre l'échantillon testé et la séquence de référence. Là encore plusieurs logiciels existent. Ils prennent en compte les scores de qualité des bases et des alignements pour distinguer les véritables variations de la séquence analysée, des erreurs générées par le séquençage ou l'alignement. Toutes les variations détectées (SNV et insertions/délétions) sont stockées dans un fichier normalisé au format ".vcf" (Variant Call Format).

Le *variant calling* peut révéler un très grand nombre de variations. Jusqu'à 20 000 pour le séquençage d'un exome par exemple. Pour pouvoir être filtrées et interprétées, toutes ces variations doivent être annotées. Les logiciels d'annotation permettent d'associer, à partir de la position génomique, des informations telles que le gène portant la variation, la position au sein du gène ou le type de variation (faux-sens, non-sens, variant d'épissage...). Ils sont également capables d'identifier les variations polymorphiques à partir de différentes bases de données rapportant la fréquence des variations dans la population générale comme le site 1000 Genomes, qui rassemble les données de séquençage du génome de 1000 individus a priori exempts de pathologie mendélienne, formant ainsi une base complète de polymorphismes (<http://www.1000genomes.org>). Le site ExAC (Exome Aggregation Consortium, <http://exac.broadinstitute.org/>) rassemble des données de séquençage de plus de 60 000 exomes. Les individus séquencés ne présentent pas de pathologie pédiatrique sévère mais pourraient être atteints de maladies multifactorielles (asthme, autisme, schizophrénie). Il est donc probable que cette base contienne des variations pathogènes.

Les logiciels d'annotation peuvent également interroger des bases de données comme OMIM (Online Mendelian Inheritance in Man, <https://www.omim.org>), qui contient des

informations sur toutes les maladies mendéliennes humaines connues et plus de 4000 gènes associés à un phénotype. D'autres bases de données listent les variations génétiques et donnent des informations sur leur degré de pathogénicité. On peut citer la base de données ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>) ou HGMD (Human Gene Mutation Database, <http://www.hgmd.cf.ac.uk>).

L'impact des variations faux-sens peut aussi être prédit par des scores de pathogénicité calculés par des outils *in silico*. Certains algorithmes sont basés sur la conservation des acides aminés au cours de l'évolution. D'autres prennent en compte les propriétés physicochimiques des acides aminés ou leur implication dans des structures particulières de la protéine comme les ponts disulfures, les cœurs hydrophobes ou les domaines fonctionnels. Ces annotations sont une aide précieuse pour filtrer, prioriser et interpréter les variations détectées.

III.1.e. Interprétation des variations

Après l'annotation des variations identifiées lors du séquençage d'un exome, plusieurs filtres sont appliqués. Tout d'abord des filtres de qualité permettent d'éliminer les artéfacts dus à des erreurs de séquençage ou à des difficultés d'alignement des lectures. Ensuite, ne sont gardées que les variations non synonymes, proches des sites d'épissage ou les petites insertions et délétions (variations NSSSI : Non Synonymous, Splice Site, Insertion/deletion) de faible fréquence dans la population générale.

Parmi les variations retenues après analyse informatique, plusieurs types de situations peuvent se présenter. Le cas le plus simple est la découverte d'une variation connue et identifiée comme pathogène dans les bases de données (OMIM, Clinvar, HGMD). Quand la variation n'est pas rapportée comme pathogène mais concerne un gène déjà impliqué dans la pathologie étudiée, sa classification est plus ou moins aisée selon le type de variation. Les insertions ou délétions de quelques bases décalant le cadre de lecture ou les variations non-sens interrompant la séquence génique ne posent généralement pas de problème d'interprétation et sont le plus souvent classées comme pathogènes. Les variations d'épissage sont également peu ambiguës, puisqu'elles perturbent à priori la production du transcrit. Leur effet délétère peut être confirmé par un séquençage ciblé des ARN messagers du gène. La principale difficulté d'interprétation réside dans l'analyse des variations faux-sens: elles correspondent à un changement d'acide aminé et peuvent avoir un effet très variable allant de l'absence de conséquence sur l'activité de la protéine jusqu'à une perte totale de sa fonction. Plusieurs outils sont utilisés pour tenter de prédire l'effet de ces substitutions, en s'appuyant sur des logiciels de prédiction de pathogénicité évoqués précédemment. Les variations faux-sens les plus à même d'avoir un effet pathogène sont celles impactant des acides aminés conservés au cours de l'évolution ou impliqués dans des domaines fonctionnels ou structuraux de la protéine. L'évaluation du caractère pathogène d'une variation s'appuie enfin sur l'étude de la ségrégation familiale. Ainsi, pour valider l'identification d'une nouvelle variation, même dans un gène connu, le mode de transmission et le lien entre le phénotype et le génotype chez plusieurs membres de la famille doivent être contrôlés.

L'utilisation de techniques de séquençage pan-génomique comme le séquençage d'exome ou de génome peut conduire à l'identification de variations non connues dans la pathologie étudiée et nécessite des étapes d'explorations complémentaires qui relèvent plus du champ de la recherche que de celui du diagnostic. Cet aspect de l'interprétation est développé dans le chapitre III.2.b.

III.2. Le séquençage de l'exome

Depuis 2010, le NGS a révolutionné le monde de la génétique médicale. Le séquençage de l'exome (ES: exome sequencing) permet de séquencer toutes les régions codantes du génome en quelques heures. Il représente environ 34 millions de paires de bases, soit 1,2% de la totalité du génome. Il existe des versions étendues de l'ES incluant les exons non codants ainsi que des régions régulatrices. Il est donc possible de détecter toute sorte de variations nucléotidiques et le développement récent d'algorithmes spécifiques permet aujourd'hui la détection de variations structurales, telles que les grandes délétions ou duplications de l'ADN (Fromer et al., 2012; Hayes et al., 2013; Krumm et al., 2012). On estime que l'exome contient 85 % des variations responsables de maladies génétiques (Botstein and Risch, 2003; Majewski et al., 2011). L'ES a donc le potentiel de mettre évidence les causes d'un grand nombre de maladies génétiques rares et d'identifier des variations prédisposant à des maladies plus communes comme les cardiopathies ou les cancers.

Il présente également un avantage considérable par rapport au séquençage de panels de gènes associés à une pathologie, à savoir la possibilité de ré-analyse des données. En effet, les analyses d'exome dont l'interprétation n'a pas identifié de variation causale peuvent être réinterprétées périodiquement à la lumière des nouvelles connaissances scientifiques. Cela n'est pas possible sur des données de panels. Il faudra alors mettre à jour la composition du panel et re-séquencer les échantillons concernés. Plusieurs équipes ont rapporté leur expérience de ré-analyse des données avec des pratiques très variées et des résultats hétérogènes, mais démontrant toujours une augmentation du taux de diagnostic moléculaire. Ce travail de ré-analyse, mené par l'équipe de génétique du CHU de Dijon sur les anomalies du développement et la déficience intellectuelle a permis de réaliser 24 (15%) diagnostics moléculaires supplémentaires sur 156 analyses d'exomes négatives après un première analyse (Nambot et al., 2018).

III.2.a. Différentes stratégies d'analyse d'exome

Après annotation et filtration des variations obtenues lors du séquençage d'un exome selon les critères définis précédemment, on aboutit à la production d'un rapport contenant entre 300 et 500 variations. Malgré l'aide des outils de prédiction de pathogénicité et de stratégies de priorisation, l'interprétation de toutes ces variations peut être extrêmement fastidieuse. Des stratégies d'analyse comparatives d'exome ont été imaginées pour réduire le nombre de variations à interpréter.

- **Stratégie des exomes en trio:** elle s'applique lorsqu'un seul individu est atteint dans une famille et que l'on suspecte l'apparition d'une mutation *de novo* dominante. Dans un premier temps, on séquence l'exome du cas index et de ses parents sains. Puis on soustrait de la liste des variations du cas index ceux qui sont hérités de l'un ou l'autre des parents pour ne conserver que les variations *de novo*. Cette stratégie peut également faciliter l'identification de variations homozygotes ou hétérozygotes composites avec un allèle hérité de chaque parent. On peut l'appliquer aux formes de cancer très précoces sans agrégation familiale.
- **Stratégie des apparentés distants:** elle consiste à comparer les résultats d'exome de plusieurs individus d'une famille atteints d'une même pathologie et à sélectionner les variations qu'ils ont en commun. Le succès de cette stratégie repose essentiellement sur le choix des individus. Plus ils sont éloignés, moins ils partagent de matériel génétique. Ainsi, l'identification du gène délétère sera plus aisée si les individus atteints et séquencés ne sont pas apparentés au premier degré. Cette stratégie s'applique aux formes familiales de cancer avec de nombreux cas dans la même branche parentale.
- **Stratégie interfamiliale:** elle consiste à comparer les résultats d'exome de cas index de différentes familles atteints de la même maladie génétique pour identifier un gène altéré en commun. Elle suppose une absence d'hétérogénéité génétique, et était la stratégie de choix avant l'arrivée des nouvelles technologies.

En plus de ces trois stratégies appliquées classiquement en génétique constitutionnelle, l'accès à du matériel tumoral peut permettre de comparer l'exome constitutionnel et somatique. Cette méthode peut permettre de détecter une perte d'hétérozygotie ou un deuxième événement somatique conduisant à l'inactivation bi-allélique d'un gène suppresseur de tumeur ou de réparation de l'ADN.

III.2.b. Validation de la pathogénicité des variations

Lors de l'interprétation des données d'exome, deux niveaux d'analyse peuvent être envisagés:

- une interprétation à **visée diagnostique**, ne s'intéressant qu'aux gènes impliqués en pathologie humaine (gènes présents dans la base de données OMIM)
- une interprétation **visant à identifier de nouveaux gènes** responsables de la pathologie étudiée. Ce type d'analyse relève plus du domaine de la recherche et nécessite des investigations supplémentaires pour démontrer l'implication du gène dans le phénotype et la pathogénicité de la variation identifiée.

Ces investigations commencent généralement par une recherche bibliographique approfondie sur la fonction du gène, les voies métaboliques et les processus biologiques dans lesquels il intervient. Il existe des sites internet tels que Gene Ontology Consortium (<http://www.geneontology.org/>), UCSC Genome Browser (<http://genome.ucsc.edu>) ou Uniprot (<https://www.uniprot.org/>) qui facilitent l'accès à ces informations. La recherche de modèles animaux mutés pour le gène en question peut également permettre de vérifier son implication dans le phénotype.

La réplication d'une association génotype-phénotype (autres patients ayant un phénotype similaire et une variation dans le même gène) est un argument fort en faveur de la pathogénicité d'une variation. Elle peut se faire par recherche bibliographique, par des échanges et présentations de cas cliniques lors de congrès nationaux ou internationaux ou par le séquençage ciblé du gène sur une cohorte de patients présentant un phénotype similaire. Mais parfois, la rareté de la pathologie ou son hétérogénéité génétique font qu'il est difficile d'identifier d'autres patients par ces méthodes. Un moyen efficace de partager les données cliniques et biologiques pour augmenter la récurrence a dû être développé. Des plateformes de soumission en ligne ont vu le jour ces dernières années permettant à chaque généticien de partager de manière anonyme les données cliniques et moléculaires de ses patients. Ce partage international de données permet d'identifier rapidement d'autres individus porteurs de variations dans des gènes candidats et de confirmer ou d'infirmer leur pathogénicité (Kernohan et al., 2017; Shaheen et al., 2016). Ces échanges rapides inter-équipe ont permis de faire exploser l'identification de nouveaux gènes responsables de maladies génétiques humaines. Ceci est d'ailleurs illustré par le fait que le rendement diagnostique en 2012-2014 était estimé à 25 % alors qu'il atteint 35% environ actuellement (Dragojlovic et al., 2018; Yang et al., 2013).

Des analyses fonctionnelles sont parfois nécessaires par la suite afin de démontrer l'impact délétère des variations génétiques suspectées comme causales sur l'expression, la localisation de la protéine, et/ou la perturbation de la voie physiologique dans laquelle la protéine intervient.

III.3. Apports scientifiques et médicaux du NGS/ES

III.3.a. Découverte de la variabilité du génome humain

Le séquençage d'exome à grande échelle a révélé une grande variabilité dans les séquences codantes avec environ 20000 SNV par exome dont 500 SNV rares avec une fréquence allélique inférieure à 0,1 % dans la population générale. Les données du consortium ExAC (ExomeAggregation Consortium), portant sur plus de 60000 exomes, montrent que 99% des variations rapportées ont une fréquence allélique inférieure à 1% et que 54% sont uniques. Elles indiquent également que chaque individu est susceptible d'être porteur d'une à deux variations codantes *de novo*, c'est-à-dire une nouvelle variation, non hérité d'un des deux parents (Acuna-Hidalgo et al., 2016; Lek et al., 2016). De plus, il a été montré un enrichissement en altérations *de novo* entraînant une perte de fonction dans différentes pathologies comme la déficience

intellectuelle ou l'autisme (Gilissen et al., 2014; Samocha et al., 2014). Toutes ces données indiquent **l'implication de variations très rares, voire privées, dans les maladies mendéliennes et posent la question de leur contribution à la prédisposition héréditaire au cancer.**

III.3.b. Apport du NGS/ES dans les maladies Mendéliennes

L'ES a initialement été utilisé dans le domaine de la recherche pour des pathologies ultra-rares, résistantes aux techniques diagnostiques traditionnelles, concernant des groupes homogènes de patients. Cette approche a démontré un taux de succès sans précédent dans l'identification de gènes associés à une pathologie humaine (Ng et al., 2010; Thauvin-Robinet et al., 2013, 2014). L'exome a ensuite été appliqué dans de grandes séries de patients présentant des maladies plus fréquentes, avec une grande hétérogénéité moléculaire et des spectres phénotypiques larges, telle que la déficience intellectuelle syndromique (Gilissen et al., 2014; Vissers et al., 2016), le retard de développement (Fitzgerald et al., 2015), l'autisme (Iossifov et al., 2014; Stessman et al., 2016), l'épilepsie (Helbig et al., 2016; Kodera et al., 2016) ou les anomalies cardiaques congénitales (Blue et al., 2014; Long et al., 2015). Une interprétation plus précise des données et une réduction des coûts de séquençage a permis son implantation dans le domaine diagnostique. Le taux diagnostique actuel de l'ES est compris entre 25 et 40% dans les anomalies du développement et la déficience intellectuelle, selon la pathologie et l'hétérogénéité de la cohorte étudiée (Iglesias et al., 2014; Retterer et al., 2016). Ce taux diagnostique correspond à l'identification d'une variation pathogène dans un gène précédemment impliqué en pathologie humaine, avec des corrélations génotype-phénotype publiées. Cependant, des preuves moléculaires ou phénotypiques peuvent manquer dans certains cas pour conclure à la pathogénicité d'une variation, aboutissant à un résultat non concluant ou négatif. Par exemple, pour une variation faux-sens détectée chez un patient avec un phénotype concordant avec la pathologie associée au gène d'intérêt, mais dans lequel seules des variations décalant ou interrompant le cadre de lecture ont été rapportées, ou dans le cas où une unique famille atteinte a été décrite dans la littérature.

III.3.c. Apport du NGS/ES en oncogénétique

La plupart des études d'ES en oncogénétique se sont intéressées à une localisation tumorale en particulier. Elles ont permis d'identifier quelques nouveaux gènes (Chubb et al., 2016; Comino-Méndez et al., 2011; Smith et al., 2013; Stafford et al., 2017) ou d'élargir le spectre des cancers associés à des gènes de prédisposition connus (Chubb et al., 2016; Powis et al., 2018; Stafford et al., 2017; Wilson et al., 2017).

La comparaison de grandes cohortes de patients et de contrôles bien définies selon des critères cliniques est une stratégie classique en cancérologie (Valle, 2014). Une étude britannique a appliqué cette méthode sur les données d'ES de 1006 cas de cancers colorectaux (CCR) familiaux diagnostiqués avant l'âge de 55 ans et 1609 contrôles sains (Chubb et al., 2016). Elle a identifié des variations rares à forte pénétrance dans 16 % des cas de CCR. Bien que la

majorité des variations concernent des gènes connus en oncogénétique, trois gènes, *POT1*, *POLE2* et *MRE11*, ont été identifiés comme des nouveaux gènes candidats à une prédisposition au CCR. Cette étude a également permis de clarifier l'architecture génétique de la prédisposition au CCR et de probablement écarter l'existence d'autres gènes de prédisposition avec des variations à forte pénétrance qui pourraient individuellement expliquer plus de 1 % des cas familiaux.

Certaines études ont permis d'identifier de nouveaux gènes de prédisposition à partir d'un nombre limité de patients. C'est le cas du gène *MAX* dans les phéochromocytomes (Comino-Méndez et al., 2011) ou de *SMARCE1* dans les méningiomes multiples (Smith et al., 2013). Dans les deux cas, l'ES de 3 patients non apparentés sans mutation causale dans les gènes de prédisposition connus pour leur cancer a permis d'identifier le gène candidat. Des répliques ont ensuite été trouvées chez d'autres patients par séquençage ciblé. Cela valide l'utilité de l'ES pour identifier des variations rares à partir de quelques patients bien sélectionnés.

La stratégie d'exome en trio a été utilisée avec succès par une équipe française en 2014. Cette méthode a été appliquée chez une patiente présentant un adénocarcinome séreux métastatique de l'ovaire à 21 ans sans antécédents familiaux. Elle a permis d'identifier une variation *de novo* du gène *INHBA* codant pour la sous unité β A des inhibines et activines, qui jouent un rôle important dans le développement ovarien. *In vitro*, cette mutation altère le ratio inhibine/activine sécrété. En réalisant le séquençage ciblé des gènes des inhibines, l'équipe a identifié une variation du gène *INBA* codant la sous unité α , partenaire de la sous unité β A. Cette variation altère également le ratio inhibine/activine en perturbant la biosynthèse des inhibines A et B. Puis une seconde variation du gène *INHBA* a été identifiée dans une cohorte de réplique de 62 cas de cancer ovarien précoce (Tournier et al., 2014).

Ces différents exemples montrent l'utilité d'approfondir l'analyse génétique par une technique pan-génomique et des méthodologies issues du domaine des maladies rares, telles que l'ES en trio, pour les cas les plus évocateurs d'une prédisposition génétique, quand les approches traditionnelles n'ont pas permis d'identifier la cause de la prédisposition.

Problématique et hypothèse de recherche

Dans les années 90, des études de liaison génétique dans des familles à haut risque ont permis d'identifier de nombreux *loci* rares à forte pénétrance, comme ceux de *BRCA1* et *BRCA2* pour les cancers du sein ou de *APC* pour les cancers colorectaux (Fletcher and Houlston, 2010; Foulkes, 2008). Les variants identifiés par ce type d'études ont une fréquence faible dans la population générale (< 1/1000) et confèrent une augmentation du risque de cancer de 5 à 50 fois par rapport à la population générale (Stadler et al., 2010).

Les études de gènes candidats, basées sur la comparaison de la fréquence des mutations dans des groupes patients et de témoins ont également permis l'identification de plusieurs GPC. Les gènes candidats ont été proposés sur la base de leur implication dans les mêmes voies métaboliques que les GPC déjà identifiés (par exemple les voies de réparation de l'ADN ou impliqués dans le raccourcissement des télomères) ou parce qu'ils sont le siège de mutations somatiques dans les tumeurs (Rahman, 2014). Les variations de la plupart des GPC identifiés par ces études ont des fréquences alléliques comprises entre 1/1000 et 1/100 et une pénétrance modérée avec un risque de cancer environ 2 fois plus fort que dans la population générale (Stadler et al., 2010).

A partir de 2005, les études d'association pan-génomique (GWAS : Genome Wide Association Studies) ont permis d'étudier l'hypothèse de l'implication de variations communes (fréquence allélique > 10%) à faible pénétrance dans différentes pathologies dont le cancer. Ces études ont permis d'identifier 391 variations communes, ne conférant qu'une faible augmentation du risque de développer un cancer (Hindorff et al., 2009). Presque tous les *loci* identifiés par les GWAS sont associés avec une faible augmentation du risque de cancer (odds ratio inférieurs à 1,5). La plupart sont situés dans des régions inter-géniques. Seulement 15 se trouvent dans des GPC connus, mais aucun n'est associé aux cancers qui surviennent chez les porteurs de mutations rares et pénétrantes de ces gènes. Ces données suggèrent que les mécanismes associés aux variations rares à forte pénétrance et aux variations fréquentes à faible pénétrance sont distincts (Rahman, 2014). Cela diffère d'autres maladies complexes communes qui montrent un chevauchement entre ces deux composantes de leur architecture génétique (Morris et al., 2012; Teslovich et al., 2010).

Comme le montre la Figure 9A, l'addition des GPC à plus ou moins forte pénétrance et des SNP à faible pénétrance identifiés par les GWAS ne représente qu'une faible proportion (entre 15 et 30 %) du risque familial des cancers les plus communs comme le cancer du sein, du colon ou de la prostate. Ces trois types d'études ont permis de définir l'architecture génétique de la prédisposition au cancer en 3 groupes : un premier groupe de gènes avec des mutations rares à forte pénétrance, un groupe intermédiaire avec des mutations plus fréquentes mais des

pénétrances plus faibles et un troisième groupe composé des SNP communs identifiés par GWAS qui ne confèrent qu'une faible augmentation du risque de cancer (Figure 9B).

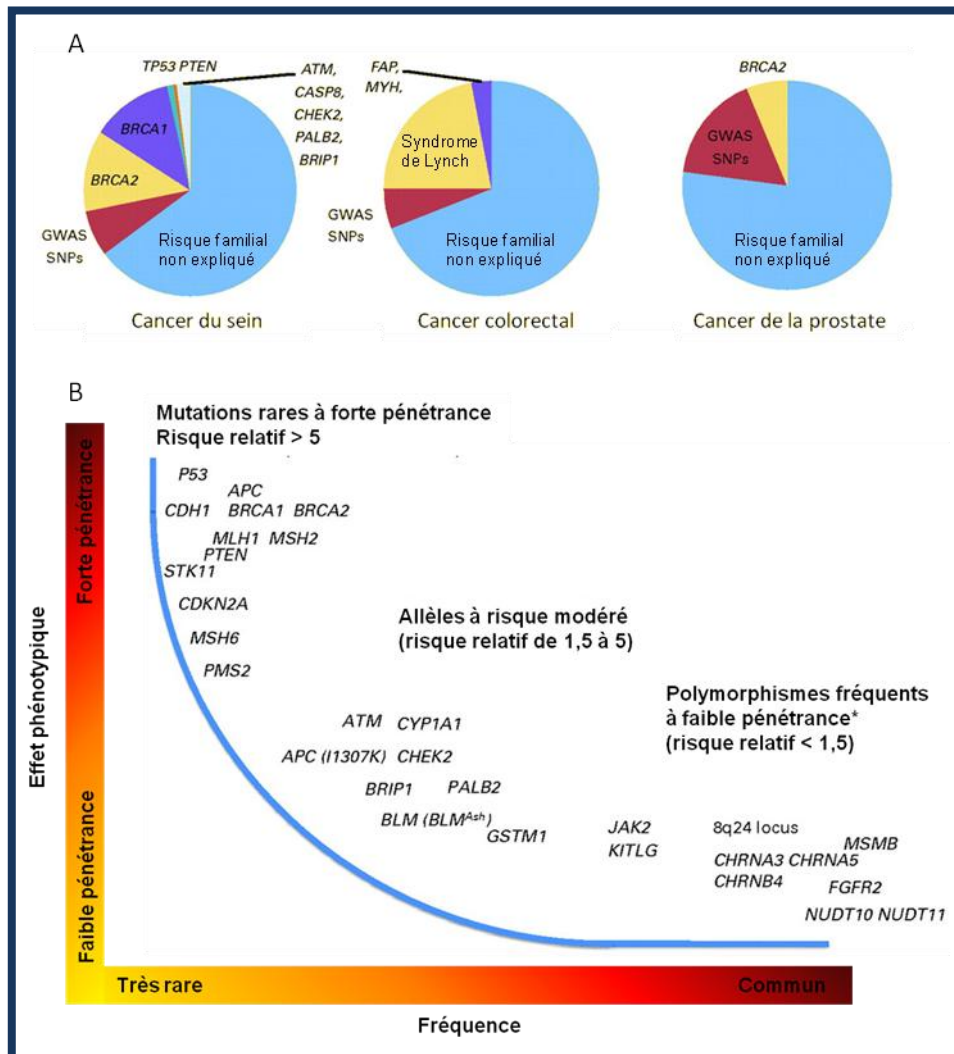


Figure 9: A: Proportion du risque familial expliqué par les GPC connus et les SNP identifiés par GWAS pour trois cancers communs. B: Effet phénotypique et fréquence des allèles de prédisposition au cancer. (*) Les noms de gènes représentent le gène candidat le plus probable à proximité du SNP identifié par GWAS. (adapté de Stadler et al., 2010).

Une autre hypothèse qui permettrait de découvrir de nouveaux gènes est celle de l'implication de variants ultra rares voir privés à forte pénétrance comme cela a été démontré dans le domaine des maladies génétique rares, grâce à l'arrivée du séquençage haut débit de l'exome. **L'objectif de ce travail est d'explorer cette hypothèse dans les prédispositions génétiques aux cancers en appliquant des stratégies ayant fait leur preuve dans le domaine des maladies rares, à partir de phénotypes extrêmes de cancer. Pour cela, deux stratégies sont proposées, le séquençage en trio dans les cas précoces sporadiques, et le séquençage d'apparentés éloignés dans les formes familiales extrêmes.**

Malgré le développement du séquençage haut débit, seuls les GPC associés au type de cancer du patient sont séquencés en cas de suspicion d'une prédisposition au cancer. Le résultat est souvent négatif et laisse bon nombre de patient sans diagnostic moléculaire. L'approfondissement des investigations génétiques par l'ES chez des patients négatifs pour les GPC classiques pourrait permettre d'augmenter le rendement diagnostique et d'élargir le spectre des cancers associés à certains gènes comme cela à déjà été montré pour les cancers du colon, du sein ou de l'ovaire (AlDubayan et al., 2018; Espenschied et al., 2017; Pearlman et al., 2017; Yurgelun et al., 2017).

Matériel et méthodes

I. Extraction de l'ADN génomique

- A partir de sang total

Les ADN génomiques (ADNg) sont extraits à partir de 3 à 5 mL de sang total à l'aide du kit Genra Puregene (Qiagen GmbH) en suivant le protocole recommandé par le fournisseur.

- A partir de tissu tumoral congelé

Vingt milligrammes de tissu congelé sont lysés pendant une nuit à 56 °C sous agitation modérée dans 180 µL de tampon ATL et 20 µL de protéinase K du kit QIAamp DNA Mini (Qiagen GmbH). L'ADN génomique est ensuite été extrait de ce lysat en suivant le protocole *DNA Purification from Tissues*.

- A partir de tissu fixé au formol et inclus en paraffine (FFPE: Formol Fixed and Paraffin Embedded)

Quinze copeaux de tissu FFPE de 20 µm d'épaisseur ont été coupés à l'aide d'un microtome. Trois kits différents sont ensuite utilisés selon les recommandations des fournisseurs pour l'extraction d'ADN : extraction automatisée sur l'automate QIASymphony (Qiagen GmbH), utilisation du kit Genra Puregene (Qiagen GmbH) ou du kit QIAamp® DNA Mini (Qiagen GmbH). Les copeaux de tissu FFPE sont digérés une nuit à 56 °C sous agitation modérée avec le tampon de lyse et la protéinase K fournis dans chaque kit. Les lysats sont ensuite incubés 1 heure à 95 °c pour casser les liaisons covalentes qui se forment lors de la fixation des tissus. La suite de l'extraction d'ADN a été faite en suivant le protocole d'extraction à partir de tissus FFPE de chaque kit.

II. Dosage et contrôle qualité de l'ADN génomique

L'ADN génomique est dosés par fluorimétrie à l'aide du kit Qubit DS DNA broad range (Thermo Fischer Scientific Inc.), et par mesure de l'absorbance à l'aide du spectrophotomètre Multiscan GO (Thermo Fischer Scientific Inc.) pour déterminer les rapports A260/A280 et A260/A230. La qualité des ADNg est vérifiée par électrophorèse sur gel d'agarose à 1% dans du tampon TBE (Triss – Borate – EDTA, Thermo Fischer Scientific Inc.).

III. Extraction, dosage et contrôle qualité des ARN totaux

Les ARN totaux sont extraits à partir de sang total prélevé en tube PAXgene (Preanalytics GmbH) à l'aide du kit PAXgene Blood RNA (Preanalytics GmbH) en suivant le protocole standard. Les ARN totaux sont ensuite dosés par mesure de l'absorbance (Multiscan GO, ThermoFischer Scientific Inc.), leur qualité est évaluée en déterminant le RNA integrity number (RIN) à l'aide du kit RNA nano 6000 et du bioanalyzer 2100 (Agilent Technologies Inc.).

IV. Séquençage d'exome (ES)

L'ES est sous traité sur la plateforme de biologie et de pathologie des tumeurs du centre anti-cancer Georges François Leclerc de Dijon (CGFL). L'ADN génomique est tout d'abord fragmenté (tailles comprises 150 et 200 pb) puis les extrémités des fragments sont réparées. Une adénine est ajoutée par ligation aux extrémités 3' des fragments d'ADN pour permettre l'ajout des séquences adaptatrices contenant une thymine libre à l'extrémité 3'. Les fragments d'ADN liés aux séquences adaptatrices sont ensuite amplifiés par PCR pour constituer des banques de fragments d'ADN génomique. Ces étapes sont réalisées avec le kit SureSelect XT Library Prep Kit ILM (Agilent Technologies, Inc.). L'étape suivante consiste à extraire de ces bibliothèques les séquences codantes, constituant l'exome. Pour cela, les banques d'ADN génomique sont hybridées avec une banque de sondes biotynylées fournie par le kit SureSelect Clinical Research Exomes (Agilent Technologies, Inc.). Les hybrides sont ensuite purifiés en utilisant des billes magnétiques couplées à la stréptavidine. Ces nouvelles banques d'ADN sont amplifiées par PCR avec des amorces contenant des séquences d'index permettant d'identifier les échantillons. Le séquençage est enfin réalisé sur le séquenceur NextSeq 500 (Illumina Inc.) en lecture paires de 120 bases.

V. Analyse des données brutes de séquençage haut débit

Les données brutes de séquençage d'exome sont obtenues au format fastq. Elles sont d'abord alignées sur le génome de référence GRCh37/hg19 grâce au logiciel Burrows-Wheeler Aligner (BWA, v 0.7.15). Les potentiels duplicats de PCR sont marqués avec le logiciel Picard (v2.4.1). La suite d'outils bioinformatiques Genome Analysis ToolKit (GATK, v 3.7) est utilisée pour recalibrer le score de qualité des bases, réaligner les lectures contenant des insertions et délétions et rechercher les variations de séquence. L'application SeattleSeq Annotation est utilisée pour annoter les variations. Un programme bioinformatique développé en interne au laboratoire permet de générer un rapport contenant uniquement les variations NSSI rares (fréquence allélique inférieure à 1% dans les bases de données dbSNP150 et GnomAD, (r2.0.2)).

VI. Séquençage haut débit ciblé

– Séquençage de gènes candidats

Les amorces utilisées pour amplifier les régions codantes des gènes à séquencer sont conçues grâce à l'application ExonPrimer (<https://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html>) (Annexe 2). Le principe étant de couvrir un maximum d'exons dans le même amplicon pour limiter le nombre de PCR à réaliser. La taille maximale des amplicons a été fixée à 9 kb. Les amplifications sont réalisées avec le kit PrimeStar GXL (Takara Bio Inc.). Après vérification des produits de PCR par électrophorèse sur gel d'agarose en TBE à 1%, les amplicons d'un même patient sont regroupés puis purifiés à l'aide de billes magnétiques AMPure XP (Beckman Coulter Inc.). Les bibliothèques pour le séquençage sont préparées à partir de ces banques d'amplicons à l'aide du kit Nextera XT (Illumina Inc.) puis séquencées sur le séquenceur haut débit MiSeq (Illumina Inc.) en lectures paires de 150 pb. L'analyse des données brutes est réalisée avec le même programme bioinformatique que pour l'analyse des données d'exome avec quelques adaptations pour du séquençage ciblé.

– Séquençage ciblé sur ADNc

Les ARN sont convertis en ADNc à l'aide du kit QuantiTect Reverse Transcription (Qiagen GmbH). L'ensemble des séquences codantes des régions d'intérêt sont extraites à partir du site UCSC genome browser (<http://genome-euro.ucsc.edu/index.html>). Les amorces sont positionnées dans les exons encadrant l'exon contenant la variation d'intérêt à l'aide de l'application BatchPrimer 3 (<https://probes.pw.usda.gov/batchprimer3/index.html>) (Annexe 2). Les PCR et le séquençage haut débit ont été réalisés selon le protocole décrit précédemment pour le séquençage des gènes candidats. Les données brutes sont alignées grâce au logiciel STAR (v2.5.2b) spécialisé pour l'alignement de données de séquençage d'ADNc. Le logiciel IGV (Integrated Genome Viewer, v2.5.2) est utilisé pour visualiser les lectures et générer les représentations graphiques.

VII. Séquençage selon la méthode de Sanger

Des amorces encadrant la variation de séquence sont conçues à l'aide de l'application BatchPrimer 3 (<https://probes.pw.usda.gov/batchprimer3/index.html>) afin d'amplifier un fragment d'ADN de 300 à 800 pb (Annexe 2). Les PCR sont réalisées à l'aide du kit HotStar Taq DNA polymerase (Qiagen GmbH) en suivant les recommandations du fabricant. Les produits de PCR sont séparés par électrophorèse sur gel d'agarose à 1 % pour vérifier la bonne amplification à la taille attendue et la pureté de l'amplicon. Cinq microlitres de produits de PCR sont ensuite purifiés grâce au kit ExoSAP-IT (USB Corporation). Les réactions de séquence sont préparées et purifiées à l'aide des kits BigDye Terminator v1.1 et BigDye XTerminator (ThermoFischer Scientific Inc.) en suivant le protocole standard. Les produits de séquençage sont ensuite analysés par électrophorèse capillaire un analyseur génétique ABI

PRISM 3130xl (ThermoFischer Scientific Inc.). Les résultats sont analysés à l'aide du logiciel Mutation Surveyor (Softgenetics).

VIII. Recherche de perte d'hétérozygotie par analyse chromosomique sur puce à ADN (ACPA)

L'ACPA comparative entre l'ADN constitutionnel et tumoral est réalisée sur une puce SurePrint G3 Human 4x180 k (Agilent Technologies Inc.) selon les recommandations du fournisseur. Les ADN génomiques sont digérés avec les enzymes de restriction AluI/RsaI pendant 2 heures à 37°C. L'ADN du patient et de la tumeur sont marqués respectivement avec les fluorochromes Cyanine5 (Cy5) ou Cyanine3 (Cy3), par amorçage aléatoire, à l'aide de l'enzyme Exo-Klenow (kit Agilent Genomic DNA labeling Kit Plus), puis purifiés sur colonnes QIAquick PCR Purification (Qiagen GmbH). Un dosage sur un Nanodrop (ThermoFischer Scientific Inc.) permet de vérifier l'efficacité du marquage. Les ADN marqués sont ensuite co-hybridés sur la lame pendant 40h à 65°C sous agitation par rotation constante. Après lavage, la lame est lue par le scanner G2565BA (Agilent Technologies Inc.). Le logiciel Feature Extraction (Agilent Technologies Inc.) est utilisé afin d'obtenir les données brutes et le logiciel Genomic WorkBench 6.5 (Agilent Technologies Inc.) pour l'analyse des données.

IX. Culture cellulaire et analyses fonctionnelles sur lignée cellulaires

IX.1. TP53I3

– Conditions de culture

La culture primaire de fibroblastes du patient a été établie à partir d'une biopsie de peau par le laboratoire de cytogénétique du CHU de Dijon. Les LCF fibroblastes témoins proviennent de biopsies de peaux de donneur sains. Les fibroblastes sont cultivées dans du milieu DMEM (Dulbecco's Modified Eagles Medium) contenant 4.5g/L de glucose (Thermo Fisher Scientific Inc.) supplémenté de 10% de sérum de veau foetal (SVF) et 1% d'un cocktail antibiotique (ZellShield, Minerva Biolabs GmbH), dans un incubateur à atmosphère humide, à 37°C et 5% de CO₂.

– Extraction des protéines

Les fibroblastes sontensemencés et cultivées une nuit en plaque 6 puits (100000 cellules par puits dans 2 mL de milieu de culture). Après rinçage au PBS froid, les cellules sont lysées sur glace dans du tampon RIPA (Cell Signaling Technology Inc.) additionné de PIC (Protease

Inhibitor Cocktail, Sigma-Aldrich Inc.) et de PMSF (Phenylmethylsulfonyl fluoride, Sigma-Aldrich Inc.). Les cellules sont grattées au fond de chaque puits. Le lysat obtenu est incubé 15 min sur glace, puis centrifugé 15 min à 13 000 RPM et 4°C. Le lysat protéique est transféré dans un tube propre, et conservé à -20°C.

– Dosage des protéines

Le dosage des protéines est effectué à l'aide d'une gamme étalon de BSA (Bovine Serum Albumine) (0, 50, 100, 200, 300, 400, 500 et 700 µg/mL) et du kit de dosage protéique BCA (Thermo Fisher Scientific Inc.). La lecture de l'absorbance (562 nm) est réalisée sur un spectrophotomètre lecteur de plaque 96 puits Multiskan Go (Thermo Fisher Scientific Inc.). Le principe de ce dosage repose sur la réduction des ions cuivrique Cu(II) en Cu(I) en milieu alcalin et sur l'acide bicinchonique, réactif colorigène, qui réagit spécifiquement avec le Cu(I) ce qui va former un complexe de couleur pourpre. L'absorbance à 562 nm est alors proportionnelle à la quantité de protéines présentes dans le milieu réactionnel.

– Electrophorèse en SDS-Page et Western Blot

Une quantité égale de protéines de chaque échantillon est dénaturée et chargée négativement par incubation 5 min à 98°C dans du tampon Laemmli 4x (Bio-Rad Laboratories, Inc.) additionné de β-mercaptoéthanol. Les protéines sont ensuite séparées par électrophorèse dans un gel d'acrylamide à 10% après migration dans un gel de concentration à 5%. La migration est effectuée dans un tampon Tris-Glycine (25mM - 192 mM) additionné de SDS à 1%, à 70 V pour le passage dans le gel de concentration et entre 100 et 140 V pour le reste de la migration. Les protéines sont ensuite transférées sur une membrane de PVDF (Polyvinylidène difluoride, Merck KGaA), préalablement activée dans du méthanol, dans un tampon contenant du Tris (6g/L) et de l'acide borique (3g/L), à 30mA pendant 2h en chambre froide. Après saturation pendant une heure dans du PBS-tween 0.05% contenant 5% de lait, les membranes sont incubées une nuit à 4°C avec l'anticorps primaire dilué au 1/1000^{ième} dans le tampon PBS - tween - lait. Les membranes sont ensuite lavées (3 lavages de 5 min en PBS-tween 0.05%) avant l'ajout de l'anticorps secondaire couplé à la peroxydase (Cell Signaling Technology Inc.), dilué au 1/5000^{ième} dans le tampon PBS - tween - lait. La révélation est faite par ajout d'un réactif chimioluminescent (Clarity Western ECL Substrate, Bio-Rad Laboratories Inc.) et la détection de chimioluminescence sur l'instrument d'imagerie ChemiDoc MP (Bio-Rad Laboratories Inc.).

– Test de survie cellulaire

Les cellules sont tout d'abord ensemencées en triplicats dans une plaque 96 puits (4000 cellules par puits, dans 100 µl de milieu de culture). Après une nuit d'incubation, le milieu de culture est changé pour un milieu contenant des doses croissantes de doxorubicine (de 0 à 3200 ng/mL). Après 48h de culture, le taux de survie cellulaire est mesuré à l'aide du kit CyQUAN XTT Cell Viability Assay (Thermo Fischer Scientific Inc.) selon les recommandations du fabricant. Le réactif XTT (2,3-Bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2H-Tetrazolium-5-

Carboxanilide) est sensible au potentiel d'oxydoréduction cellulaire. Les cellules vivantes convertissent le XTT en un produit coloré orange. L'absorbance à 450 nm est proportionnelle à la quantité des cellules vivantes. La lecture d'absorbance est réalisée sur le spectrophotomètre Multiscan GO (Thermo Fischer Scientific Inc.). Le taux de survie cellulaire est calculé par rapport à l'absorbance moyenne des puits contenant des cellules non traitées à la doxorubicine.

IX.2. ATR

Les analyses par western blot sur les lignées cellulaires pour le gène *ATR* ont été réalisées par le laboratoire *Human DNA Damage Response Disorders* (HDDR) de l'université du Sussex selon un protocole similaire à celui décrit précédemment. Les lignées cellulaires lymphoblastiques (LCL) ont été établies à partir d'un prélèvement sanguin en tube hépariné par le laboratoire *European Collection of Authenticated Cell Cultures* avant d'être transférées au laboratoire HDDRD où elles ont été maintenues dans du milieu RPMI 1640 avec 15% de SVF additionné d'antibiotiques et de L-glutamine.

- Stimulation des cellules

Pour l'analyse de la phosphorylation de CHEK1 après induction d'un stress génotoxique, les cellules sont traitées pendant 2 heures avec 500 μ M d'hydroxyurée avant l'extraction des protéines.

IX.3. Anticorps

Les différents anticorps primaires utilisés pour les analyses par western blot sont listés dans le tableau ci-dessous.

Cible	Marque	Référence	Espèce
ATR	Santa Cruz Biotechnology	SC-515173	Souris
CHK1	Cell Signalling Technology	2G1D5	Lapin
P-Ser345-CHK1	Cell Signalling Technology	133D3	Lapin
TP53I3	GeneTex	GTX111926	Lapin

Résultats

Vingt-sept patients présentant un phénotype extrême de cancer identifiés lors des consultations d'oncogénétique et validés en réunion de concertation pluridisciplinaire (RCP) ont été inclus dans ce projet dans le cadre de l'étude EXTRICAN.

Les critères d'inclusion sont les suivants:

- Agrégation familiale de cancers du même spectre
- Cas sporadique à un âge particulièrement précoce (cancer digestif avant 40 ans, cancer ovarien avant 40 ans, cancer du sein avant 30 ans)
- Multiples tumeurs chez un même patient

La stratégie d'analyse employée dépend de la présentation familiale: ES en trio pour les sporadiques précoces ou multiples et ES cas index ou avec apparenté atteint éloigné pour les formes avec agrégation familiale. Dans certains cas, l'ES a aussi été réalisé sur l'ADN tumoral. L'analyse combinée de l'ES tumoral et constitutionnel peut permettre de faciliter l'interprétation en mettant en évidence une perte d'hétérozygotie ou un second événement somatique. Le Tableau 2 récapitule les patients inclus dans ce projet ainsi que la stratégie d'analyse employée. Les résultats de chaque patient sont analysés individuellement et en sous-groupe de pathologie. Par exemple, nous avons identifié 3 familles avec plusieurs cas de cancer du sein masculin (CSM). Le CSM étant une maladie très rare (moins de 1 % des cas de cancer du sein et moins de 1% des cancers masculins), la présence de plusieurs cas au sein d'une même famille fait suspecter une prédisposition génétique.

L'analyse des données brutes de ces exomes a révélé en moyenne 258 variations NSSI par individu avec une fréquence allélique inférieure à 1/1000 dans la population générale, étant à la recherche d'un facteur de prédisposition à forte pénétrance.

La stratégie d'interprétation a été la suivante :

- Dans un premier temps, l'interprétation a été limitée aux gènes connus dans OMIM pour leur implication en cancérologie ou responsables de syndromes associés à une prédisposition au cancer.
- Puis une liste de 661 GPC ou gènes "drivers" dans le développement des cancers a été établie à partir des listes de GPC publiées et synthétisées dans le tableau 1, de la base de données COSMIC Census (Futreal et al., 2004), et de diverses publications identifiant des GPC candidats.
- Si aucune variation n'a été retenue à partir de ces listes, l'analyse a été étendue aux autres gènes. L'utilisation des GO terms a permis d'identifier rapidement les gènes participant à des processus cellulaires impliqués dans le développement de cancers.

La priorité a été donné aux variations tronquantes (insertions et délétions hors phase et substitutions non-sens) ou ayant un impact potentiel sur l'épissage. L'interprétation des variations faux-sens étant plus délicate, d'autres paramètres tels que les scores de prédiction de leur pathogénicité, la conservation inter espèce de l'acide aminé concerné et leur implication dans des domaines fonctionnels ou structuraux de la protéine ont été pris en compte.

Ces différentes stratégies d'analyse ont permis d'identifier des variations dans deux gènes déjà connus en oncogénétique et 7 nouveaux gènes candidats dont un pour chacun des trois cas de CSM inclus dans ce projet: *ATR*, *TP53I3* et *DLC1* (Tableau 3).

Afin de mieux caractériser l'implication de certains gènes candidats, nous avons recherché d'autres variations de ces gènes par séquençage ciblé sur une cohorte de réplication de CSM et évalué leur pathogénicité par des expériences de biologie moléculaire ou cellulaire en fonction du matériel auquel nous avons pu avoir accès.

Tableau 2: Patients inclus et stratégie d'analyse d'exome

ID exome	Stratégie	Organe/pathologie	Age au diagnostique	Gène candidat
dijex664	Solo	CSM	43 ans	<i>ATR</i>
dijex1631	Solo	CSM bilatéral	61 - 80 ans	<i>TP53I3</i>
dijex1477	Solo	CSM bilatéral Prostate	54 -56 ans 70 ans	<i>DLC1</i>
dijex1632	Solo	Ovaire	35 ans	<i>DFFA</i>
dijex1377	CI + T	Ovaire	28 ans	-
dijex1630	Solo	Ovaire	30 ans	<i>BLM</i>
dijex1522	Trio	Ovaire	35 ans	-
dijex1376	Trio	Ovaire	29 ans	-
dijex1799	Solo	Ovaire	28 ans	-
dijex1800	Solo	Ovaire	28 ans	-
dijex1528	CI + T	Colon	31 ans	-
dijex1525	Trio	Colon	<30 ans	-
dijex1614	Trio	Colon - Polypose	27 ans	-
dijex2027	CI + T	Colon	37 ans	-
dijex2028	Solo	Colon	35 ans	-
dijex1616	Solo	Colon - Rectum - Adénomes tubuleux de bas grade	16 ans	<i>BIK</i>
dijex1629	Solo	Sein	30 ans	<i>FAT4</i>
dijex1801	Solo	Sein	23 ans	-
dijex2022	CI + T	Sein	29 ans	-
dijex1802	Solo	Sein	34 ans (agrégation familiale)	-
dijex2030	Solo	Sein	21 ans	-
dijex2019	Trio + T	Poumon	26 ans	<i>PALB2</i>

ID exome	Stratégie	Organe/pathologie	Age au diagnostique	Gène candidat
dijex790	Trio	Poumon	20 ans	-
dijex1803	Solo	Cerveau	29 ans	-
dijex2026	Solo	Cerveau - Glioblastome	40 ans	-
dijex2031	Trio	Cerveau	22 ans	-
dijex1999	Trio	Tumeurs multiples	7 - 37 ans	<i>SETD2</i>

CSM: Cancer du sein masculin, CI: Cas Index, T: Tumeur. Trio : Analyse comparative de l'exome du cas index et des ses deux parents pour rechercher des variations *de novo* chez le cas index. Solo : Analyse de l'exome du cas index seul.

Tableau 3: Gènes candidats identifiés par séquençage d'exome

N° exome	Organe	Gène	Syndrome/ Processus cellulaire	Variation	Fréquences	
					ExAC	gnomAD
Gènes connus comme étant responsables de prédisposition génétique au cancer						
dijex1630	Ovaire	<i>BLM</i>	Syndrome de Bloom Réparation ADN	chr15:g.91304245C>T NM000057.2: c.1642C>T - p.Gln548*	1,8E-04	1,7E-04
dijex2019	Poumon	<i>PALB2</i>	Anémie Fanconi Prédisposition au cancer du sein et du pancréas Réparation ADN	chr16:g.23647641T>C NM024675.3: c.226A>G - p.Ile76Val	2,1E-05	3,5E-05
Nouveaux gènes candidats						
dijex1616	Polypes colon	<i>BIK</i>	Apoptose	chr22:g.43524493_43524506delTGCTATAGCCTGG NM_001197.4: c.261-9_265del	Absent	3,3E-05
dijex1632	Ovaire	<i>DFFA</i>	Apoptose	chr1:g.10523119delA NM_004401.2: c.779delT - p.Leu260Trpfs*16	Absent	Absent
dijex1631	CSM bilatéral 61-80 ans	<i>TP53I3</i>	Apoptose Contrôle cycle cellulaire Réparation de l'ADN	chr2:g.24307151_24307152delAG NM_147184.1: c.45_46delCT - p.Tyr16Argfs*13	Absent	Absent
dijex664	CSM	<i>ATR</i>	Contrôle cycle cellulaire Réparation ADN	chr3:g.142217574A>C NM_001184.3: c.5423T>G - p.Leu1808*	Absent	Absent
dijex1477	CSM	<i>DLC1</i>	Croissance cellulaire	chr8:g.12957110G>C NM_182643.2: c.2736C>G - p.His912Gln	Absent	Absent
dijex1629	Sein	<i>FAT4</i>	Croissance cellulaire	chr4:g.126239769C>T NM_024582.4: c.2203C>T - p.Arg735Trp	8,3E-06	4,3E-06
				chr4:g.126372168G>A NM_024582.4: c.9997G>A - p.Val3333Ile	8,3E-06	Absent
dijex1999	Tumeurs multiples	<i>SETD2</i>	Remodelage de la chromatine	chr3:g.47142990G>A NM_014159.6: c.4973C>T - p.Ser1658Leu	Absent	Absent

Les coordonnées génomiques sont données à partir du génome de référence GRCh37/hg19. Dans le cas où plusieurs transcrits existent, le transcrit choisi pour les annotations des variations sur le cDNA ou la protéine est celui le plus exprimé dans le tissu concerné ou le transcrit de référence donné par le UCSC genome browser (<http://genome-euro.ucsc.edu/index.html>).

I. Variations identifiées dans les cancers du sein masculins

Le cancer du sein masculin (CSM) est une maladie rare. Il représente moins de 1% de tous les cancers du sein et moins de 1% de tous les cancers chez les hommes. L'incidence annuelle est estimée à environ 1 pour 100 000 hommes dans le monde et le risque est inférieur à 1 sur 1000 tout au long de la vie (Ly et al., 2013). Les taux d'incidence du CSM augmentent linéairement et régulièrement avec l'âge, l'âge moyen au diagnostic étant entre 60 et 70 ans (Korde et al., 2010). Les antécédents familiaux de cancer du sein sont un facteur de risque important, ce qui suggère l'importance des facteurs génétiques dans la susceptibilité au CSM (Basham et al., 2001; Ottini et al., 2009). Comme pour le cancer du sein féminin (CSF), *BRCA1* et *BRCA2* sont les deux gènes majeurs de prédisposition au CSM. En dehors des populations présentant des mutations fondatrices, ils représentent environ 10% des CSM (Rizzolo et al., 2013). On estime que le risque à vie de développer un CSM est compris entre 1 et 5% pour les porteurs d'une mutation de *BRCA1* et entre 5 et 10% pour les porteurs d'une mutation de *BRCA2*, comparé à un risque de 0,1% dans la population générale (Breast Cancer Linkage Consortium, 1999; Evans et al., 2010; Tai et al., 2007; Thompson and Easton, 2002). Le CSM est reconnue comme étant un cancer hormono-dépendant et induit par les œstrogènes (Brinton et al., 2015). Il est similaire au CSF post ménopause, sans perte d'expression des récepteurs aux œstrogènes et à la progestérone (ER+, PR+). Cependant, comparé au CSF, le CSM est diagnostiqué plus tardivement, à un stade plus avancé et présent un grade histologique plus bas et une plus grande proportion de tumeurs ER+ et PR+ (Anderson et al., 2010).

I.1. ATR (Ataxia Telangiectasia and RAD3 related)

L'ES du premier patient, atteint d'un CSM sans antécédents familiaux, a révélé une variation non-sens hétérozygote du gène *ATR* (c.5423T>G - p.Leu1808*), non retrouvée dans les bases de données. *ATR* a un rôle central dans les voies de réparation de l'ADN et de contrôle du cycle cellulaire. Il forme un complexe avec *ATRIP* puis se fixe sur les sites d'ADN simple brin qui apparaissent lors du blocage des fourches de réplication ou la réparation des cassures double brin. *ATR* va ensuite phosphoryler de nombreux substrats comme *CHEK1*, *TP53* ou *BRCA1* pour bloquer le cycle cellulaire et laisser le temps à la cellule de réparer les lésions de l'ADN et/ou terminer la réplication proprement (Brown, 2003; Byun, 2005; Garcia-Muse and Boulton, 2005; O'Driscoll and Jeggo, 2006; Zhang et al., 2016). *ATR* est également indispensable pour maintenir la stabilité des sites fragiles du génome et participe à la régulation de la duplication des centrosomes (Casper et al., 2002). En l'absence d'*ATR*, il peut subsister des régions d'ADN dont la réplication est incomplète. La réparation des cassures double brin ne peut plus se faire par recombinaison homologue mais par d'autres mécanismes non conservatifs aboutissant à la perte d'information génétique et des échanges de chromatines, translocations et délétions apparaissent au niveau des sites fragiles (Casper et al., 2002; Mazouzi et al., 2014; Shimada et al., 2009). L'instabilité génomique qui en résulte peut être à l'origine de la transformation

tumorale. Chez la souris, la perte d'un allèle d'ATR entraîne une augmentation de l'incidence des tumeurs (Brown and Baltimore, 2000).

Sur le plan somatique, des mutations dans une séquence répétée de l'exon 10 d'ATR ont été rapportées dans des cancer de l'endomètre et de l'estomac avec instabilité des microsatellites (Lewis et al., 2005; Menoyo et al., 2001; Vassileva et al., 2002) ainsi que des délétions de plusieurs exons dans des lignées cellulaires de lymphomes (Liu et al., 2005, 2008). Ces deux dernières études montrent également que la réponse aux dommages de l'ADN est altérée dans les cellules cancéreuses mutées pour ATR, suggérant un rôle important d'ATR dans la cancérogénèse.

I.1.a. Séquençage ciblé sur la cohorte de réplication

La cohorte de réplication est composée de 86 hommes atteints de CSM et de 28 femmes atteintes apparentées à un homme atteint. Le recrutement a été effectué localement au Centre Hospitalier Universitaire ou au centre anti cancer Georges François Leclerc (CGFL) de Dijon et par appels à collaboration national (Nancy, Clermont-Ferrand, Montpellier, Strasbourg, Reims). La cohorte est amendée par 47 femmes suivies localement et classées à très haut risque de cancer du sein par le logiciel Boadicea (Lee et al., 2014) selon leur antécédents personnels et familiaux de cancer. Tous les patients sont négatifs pour *BRCA1* et *BRCA2*.

Le séquençage haut débit ciblé d'ATR sur l'ensemble de la cohorte de réplication a révélé 7 variations. La plupart (5/7) sont des variations faux-sens. Elles sont toutes présentes dans les bases de données ExAC et gnomAD (tableau 4). La variation c.4592A>G est prédite comme délétère par les algorithmes SIFT et polyphen mais elle ne se situe pas un domaine fonctionnel ou une structure particulière de la protéine. La variation c.7303A>G se situe dans le domaine kinase d'ATR mais ses scores de prédiction sont faibles. Les deux autres variations d'ATR se situent proches de sites d'épissage. L'une (c.6221+3G>A) est présente dans les bases de données et n'est pas prédite comme affectant l'épissage par l'outil de prédiction bioinformatique Human Splice Finder (HSF) (Desmet et al., 2009). La seconde (c.7762-2A>C) est absente d'ExAC et gnomAD et est prédite comme ayant un très probablement un impact sur l'épissage. Nous avons décidé de ne retenir que cette dernière variation pour la suite de ce travail. Elle concerne un homme ayant présenté un CSM à l'âge de 71 ans. Ses antécédents familiaux ne sont pas connus.

Tableau 4: Variations d'ATR identifiées dans la cohorte de répliation

Variation: génomique cDNA (NM_001184.3) protéine	Fréquences: ExAC gnomAD	Score de prédiction in silico - Prédiction d'impact sur l'épissage					Domaine - Structure
chr3:g.142168446T>G c.7762-2A>C -	Absent Absent	HSF: Altération du site accepteur Très probablement affectant l'épissage					-
chr3:g.142203979C>T c.6221+3G>A -	2,72E-04 3,13E-04	HSF : Pas de motif d'épissage identifié Probablement pas d'impact sur l'épissage					FAT
chr3:g.142261533T>C c.3424A>G p.Ser1142Gly	5,64E-04 5,53E-04	CADD	polyPhen	GERP	Grantham	SIFT	-
		14.79	0.044	5.470	56	Toléré	
chr3:g.142272098A>G c.2776T>C p.Phe926Leu	1,21E-03 1,06E-03	CADD	polyPhen	GERP	Grantham	SIFT	-
		20.30	0.355	5.350	22	Toléré	
chr3:g.142232392T>C c.4592A>G p.His1531Arg	1,48E-04 1,27E-04	CADD	polyPhen	GERP	Grantham	SIFT	-
		22.20	1	4.890	29	Délétère	
chr3:g.142272170A>G c.2704T>C p.Ser902Pro	6,85E-04 6,08E-04	CADD	polyPhen	GERP	Grantham	SIFT	-
		12.88	0.006	5.530	74	Toléré	
chr3:g.142178115T>C c.7303A>G p.Ile2435Val	2,97E-04 2,50E-04	CADD	polyPhen	GERP	Grantham	SIFT	Kinase
		0.20	0	2.680	29	Toléré	

I.1.b. Identification d'une 3^{ème} variation d'ATR

ATR à été inclus dans le panel de gènes de prédisposition au cancer séquencé par le laboratoire de génétique moléculaire du CGFL. Cela a permis d'identifier une troisième variation chez une femme présentant un cancer du sein à 41 ans et dont la sœur et la mère ont également eu un cancer du sein à 46 ans et 60 ans respectivement. Cette variation touche le codon d'initiation de la traduction de l'ARN messager (c.1A>G – p.Met1?). Elle est absente d'ExAC et présente qu'une seule fois dans gnomAD sur 242657 allèles (fréquence: 4,03.10⁻⁶). Les mutations du codon d'initiation entraînent généralement l'absence de formation d'une protéine. Cependant, il arrive que la traduction puisse tout de même démarrer à partir d'un codon d'initiation suivant (Parsons et al., 2015). Dans le cas d'ATR, deux autres codons d'initiation se trouvent à proximité. Le premier se situe à 11 bases du codon d'initiation normal et n'est pas dans la même phase de lecture. L'initiation de la traduction à partir de ce codon, conduirait à la synthèse d'un très court polypeptide de seulement 7 acides aminés. Un second codon d'initiation alternatif se trouvant dans la même phase de lecture que le codon normal est présent 30 bases plus loin. La traduction à partir de ce dernier conduirait à la perte des 10 premiers acides aminés d'ATR (Figure 10).

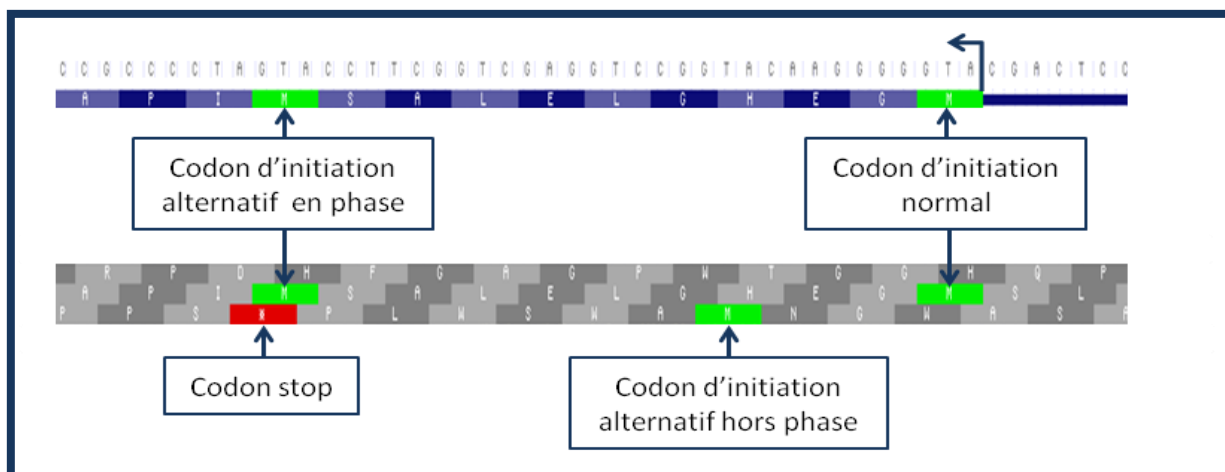


Figure 10: Troisième variation du gène ATR (c.1A>G – p.Met1?). Potentiels codons d'initiation alternatifs.

Au total, nous avons identifié 3 variations potentiellement pathogènes du gène ATR. Le tableau suivant résume ces 3 variations et le matériel auquel nous avons pu avoir accès pour essayer de démontrer la pathogénicité de ces variations et leur implication dans une prédisposition au cancer.

Tableau 5: Résumé des 3 variations du gène ATR et du matériel disponible pour chaque patient

Patient	Variation	Matériel disponible
Patient 1	c.5423T>G p.Leu1808*	ADN constitutionnel ADN tumoral Lignée cellulaire Lymphoblastique
Patient 2	c.7762-2A>C	ADN constitutionnel ARN
Patient 3	c.1A>G p.Met1?	ADN constitutionnel Lignée cellulaire Lymphoblastique Culture primaire de fibroblastes

I.1.c. Recherche de perte d'hétérozygotie dans la tumeur

Les mutations constitutionnelles hétérozygotes de gènes suppresseurs de tumeur sont souvent accompagnées d'un second évènement somatique sur l'allèle sauvage entraînant la perte totale de la fonction du gène. Ce phénomène appelé perte d'hétérozygotie (LOH: Loss Of Heterozygosity) est souvent dû à des grandes délétions. Nous avons recherché une éventuelle LOH dans la tumeur du patient 1 par une analyse chromosomique sur puce à ADN (ACPA) comparative entre l'ADN constitutionnel et l'ADN tumorale du patient. L'ADN tumoral a été extrait à partir d'un fragment de tissu tumoral congelé. La proportion de cellules tumorales a été évaluée à 80 % sur des coupes histologiques colorées en HES (hématoxyline, éosine, safran) lues

par un pathologiste. Parmi les 180 000 cibles de la puce ADN, 13 sont situées sur le gène *ATR*. L'ACPA n'a pas révélé de délétion du locus d'*ATR* (Figure 11).

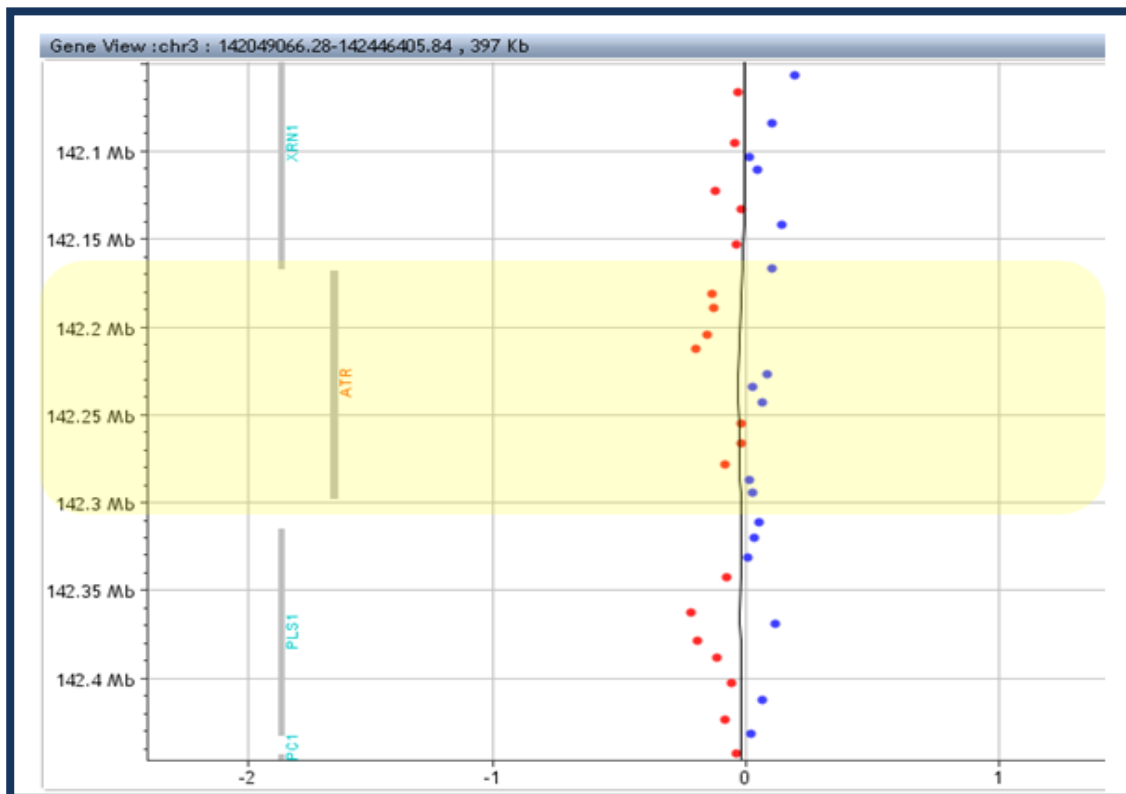


Figure 11: ACPA comparative ADN tumoral/ADN constitutionnel au locus d'*ATR*. Les points rouges et bleus dans le cadre jaune représentent les sondes couvrant le gène *ATR*.

I.1.d. Impact sur l'épissage

La variation du patient 2, identifiée dans la cohorte de réplication, touche le site accepteur d'épissage en 5' du dernier exon d'*ATR*. Elle pourrait entraîner la perte de cet exon ou un décalage du cadre de lecture et l'apparition d'un codon stop prématuré. En général, les événements touchant le dernier exon ne font pas intervenir le NMD et ne conduisent pas à la perte de la protéine mais à la présence d'une protéine tronquée de son extrémité C-terminale (Lejeune and Maquat, 2005). Les conséquences sur son activité peuvent être minimales. Cependant, le dernier exon d'*ATR* contient une partie du domaine PRD et le domaine FATC (Figure 12A). Ces deux domaines sont importants pour l'interaction d'*ATR* avec TOPBP1 et pour la bonne conformation du domaine kinase (Lempiäinen and Halazonetis, 2009). La perte de ces domaines pourrait être délétère pour la fonction d'*ATR*.

Afin d'évaluer l'impact de cette mutation sur l'épissage des ARN messagers d'*ATR*, nous avons extrait l'ARN du patient et mis au point une PCR couvrant la jonction entre les deux derniers exon d'*ATR*. Le séquençage du produit de PCR a révélé la perte des 6 premières bases du dernier exon dans environ 30% des transcrits (Figure 12B-C). Cette délétion est en phase et ne conduit qu'à la perte des deux premiers acides aminés du dernier exon d'*ATR*. Il n'y a pas de

décalage du cadre de lecture, ni de codon stop prématuré. Il est donc difficile d'évaluer l'impact de cette mutation sur la fonction d'ATR.

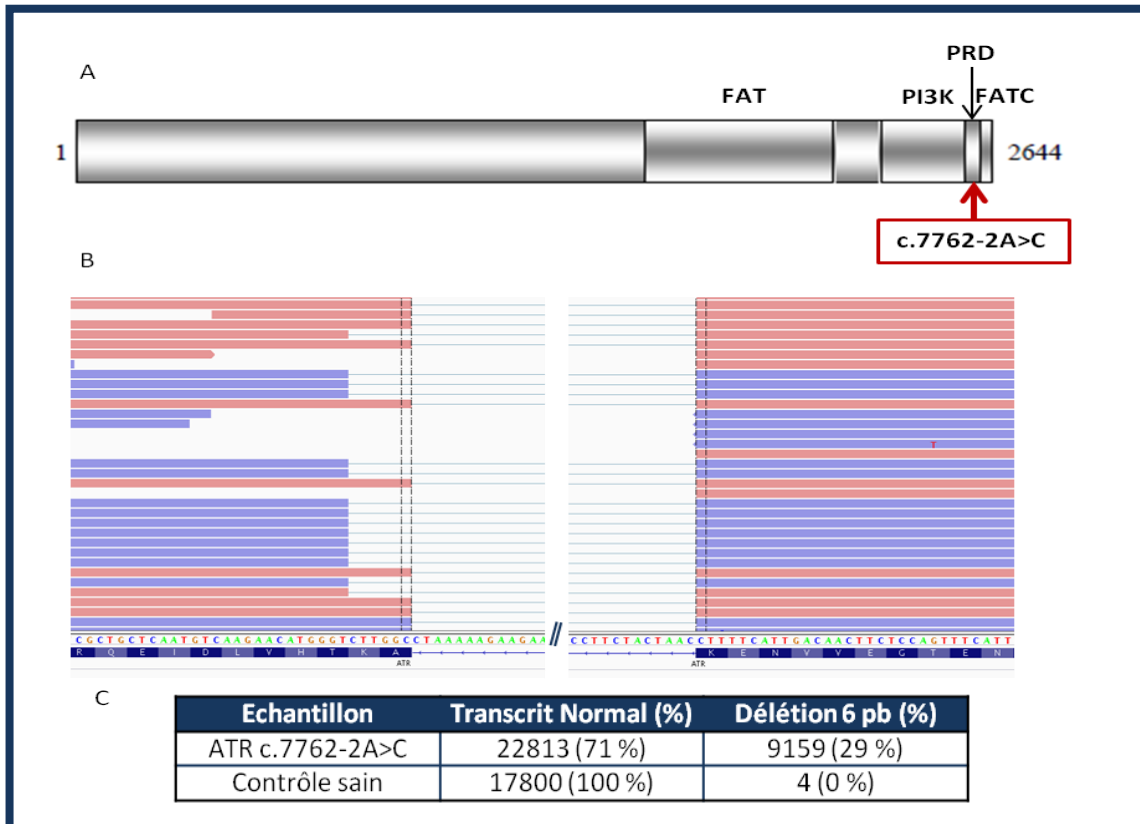


Figure 12: Impact de la variation c.7762-2A>C sur l'épissage. A: Schéma des domaines fonctionnels de l'extrémité c-terminale d'ATR. B: Visualisation sur le logiciel IGV (Integrated Genome Viewer) des lectures du séquençage des deux derniers exons d'ATR sur le cDNA du patient porteur de la variation c.7762-2A>C. C: Comptage des lectures correspondant au transcrit normal et au transcrit alternatif d'ATR sur le cDNA du patient et un contrôle non muté.

I.1.e. Expression d'ATR et phosphorylation de CHEK1

Nous avons pu établir des lignées cellulaires pour les patients 1 et 3. L'analyse d'ATR par Western blot à partir des lignées cellulaires lymphoblastiques (LCL) des patients montre une diminution du taux d'ATR d'environ 50 % par rapport à une LCL contrôle non mutée, et ne met pas en évidence de forme tronquée d'ATR qui pourrait avoir un effet dominant négatif. Nous avons également étudié la phosphorylation de CHEK1 dans la LCL du patient 1. CHEK1 est un des principaux substrats d'ATR impliqué dans le contrôle du cycle cellulaire en cas de stress génotoxique. Pour cela, les cellules ont été cultivées en présence d'hydroxyurée pour induire des cassures doubles brin dans l'ADN. Le Western blot montre une diminution d'environ 50 % de la phosphorylation de CHEK1 (pS345-CHEK1) en présence d'hydroxyurée dans les cellules du patient par rapport aux cellules contrôles (Figure 13).

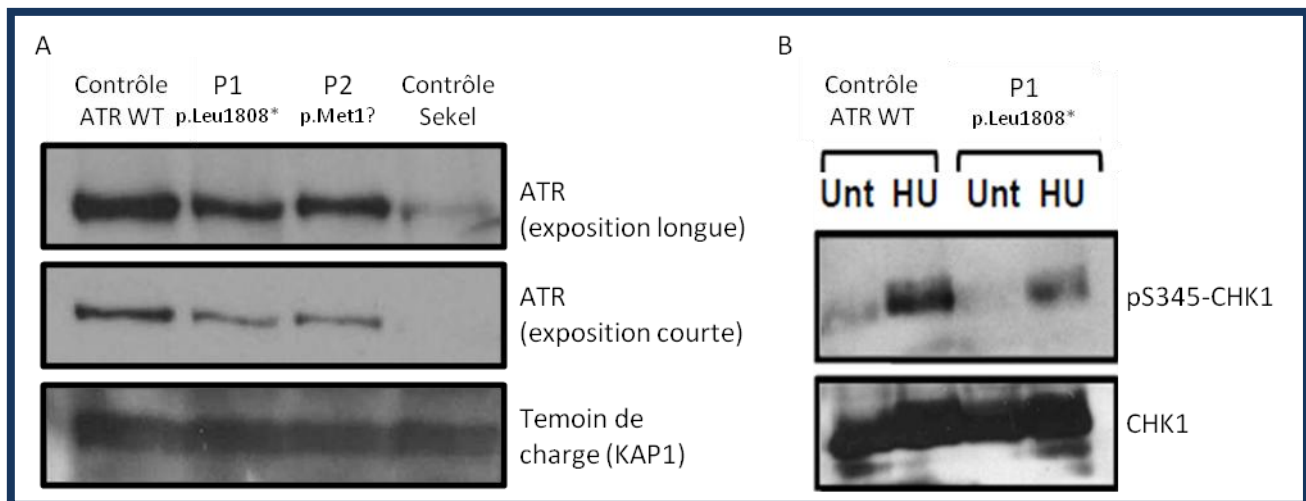


Figure 13: A : Analyse du taux d'ATR dans les LCL des patients 1 et 3. B: phosphorylation de CHK1 dans la LCL du patient 1.

Ces résultats, conjointement aux données publiées dans la littérature, seront discutés dans la section discussion, mais semblent donner des arguments supplémentaires en faveur de l'implication de ce gène dans les prédispositions génétiques au cancer du sein. Ces résultats ont déjà fait l'objet de deux présentations en congrès, dont une orale, et une à l'étranger, et un article en cours de rédaction.

I.2. TP53I3 (Tumor Protein P53 Inducible protein 3)

La variation identifiée dans le gène *TP53I3* concerne un patient atteint d'un cancer du sein bilatéral (61 et 80 ans) et d'un cancer de la prostate à 70 ans. Il s'agit d'une délétion hétérozygote de deux bases entraînant un décalage du cadre de lecture et l'apparition d'un codon stop prématuré (c.45_46delCT - p.Tyr16Argfs*13), absente des bases de données (Tableau 3). Le frère du cas index a également eu un cancer du sein à 67 ans et un cancer du colon à 80 ans. Son frère et son fils ont tout deux eu un cancer de la prostate.

TP53I3 a été identifié pour la première fois dans une étude visant à caractériser les gènes dont l'expression est augmentée lors de l'apoptose déclenchée par P53 (Polyak et al., 1997). Parmi les 14 gènes identifiés, *TP53I3* est un de ceux dont l'expression est la plus fortement augmentée. Cette étude et d'autres (Jiang et al., 2016; Polyak et al., 1997; Porté et al., 2009) montrent que *TP53I3* favorise la production d'espèces réactives de l'oxygène (ROS : Reactive Oxygen Species) dans les cellules. Les ROS ont un rôle important dans l'apoptose, d'une part en générant un stress oxydatif responsable de la dégradation de macromolécules, d'autre part en participant à la régulation des principales voies de l'apoptose (Redza-Dutordoir and Averill-Bates, 2016).

Une étude récente sur des glioblastomes montre une association entre l'expression de *TP53I3* dans les tumeurs et le grade de la maladie ainsi qu'une meilleure survie des patients ayant une forte expression de *TP53I3*. De plus, des expériences de gain et perte de fonction montrent une augmentation des capacités d'invasion et de prolifération des cellules de

glioblastome déficientes en TP53I3 alors que sa surexpression inhibe ces capacités (Quan et al., 2017).

En plus de son rôle dans l'apoptose, une étude a montré que TP53I3 colocalise avec plusieurs membres des voies de réparation de l'ADN et participe à leur recrutement aux sites de lésion de l'ADN. Sa perte d'expression diminue la phosphorylation de l'histone H2AX et peut supprimer les points de contrôles intra phase S et G2/M du cycle cellulaire (Lee et al., 2010). Cela suggère une possible implication de *TP53I3* dans les processus de contrôle du cycle cellulaire et de réparation de l'ADN.

I.2.a. Recherche de réplication

Aucune variation supplémentaire de TP53I3 n'a été identifiée dans la cohorte de réplication. Nous n'avons donc qu'un seul patient avec une variation tronquante hétérozygote. Toutefois, une étude d'ES chez 48 patientes avec cancer de l'ovaire issues de familles à haut risque de cancer du sein et de l'ovaire et négative pour BRCA1/2 a rapporté deux patientes avec une variation non-sens de TP53I3 (c.755C>G, p.Ser252*) (Stafford et al., 2017). La fréquence de cette variation dans les bases de données Exac et gnomAD est respectivement de 0,00093 et 0,00098.

I.2.b. Ségrégation familiale de la variation

D'autres cas de cancer sont présents dans la famille. Il est donc intéressant de vérifier la ségrégation de la variation dans la famille. Malheureusement, nous n'avons eu accès qu'à des blocs de tissu tumoraux fixés au formol (FFPE: Formol Fixed and Parafin Embended) pour le frère et le fils du cas index. Les ADN qui en ont été extrait sont de très mauvaise qualité. Il n'a pas été possible de réaliser le séquençage sur l'ADN du frère malgré l'essai de plusieurs kits d'extraction d'ADN, de plusieurs couples d'amorces avec des amplicons de très petite taille et de différentes polymérases.

Le séquençage Sanger a fonctionné sur l'ADN de la tumeur de la prostate du fils diagnostiqué à l'âge de 57 ans et a révélé l'absence de la variation de *TP53I3* (Figure 14).

I.2.c. Recherche de perte d'hétérozygotie

La faible qualité de l'ADN extrait à partir de la tumeur du sein FFPE du cas index n'a pas permis de réaliser une ACPA pour rechercher la perte d'hétérozygotie. Le séquençage Sanger montre que la variation est hétérozygote dans la tumeur donc il ne semble pas y avoir de perte d'hétérozygotie (Figure 14).

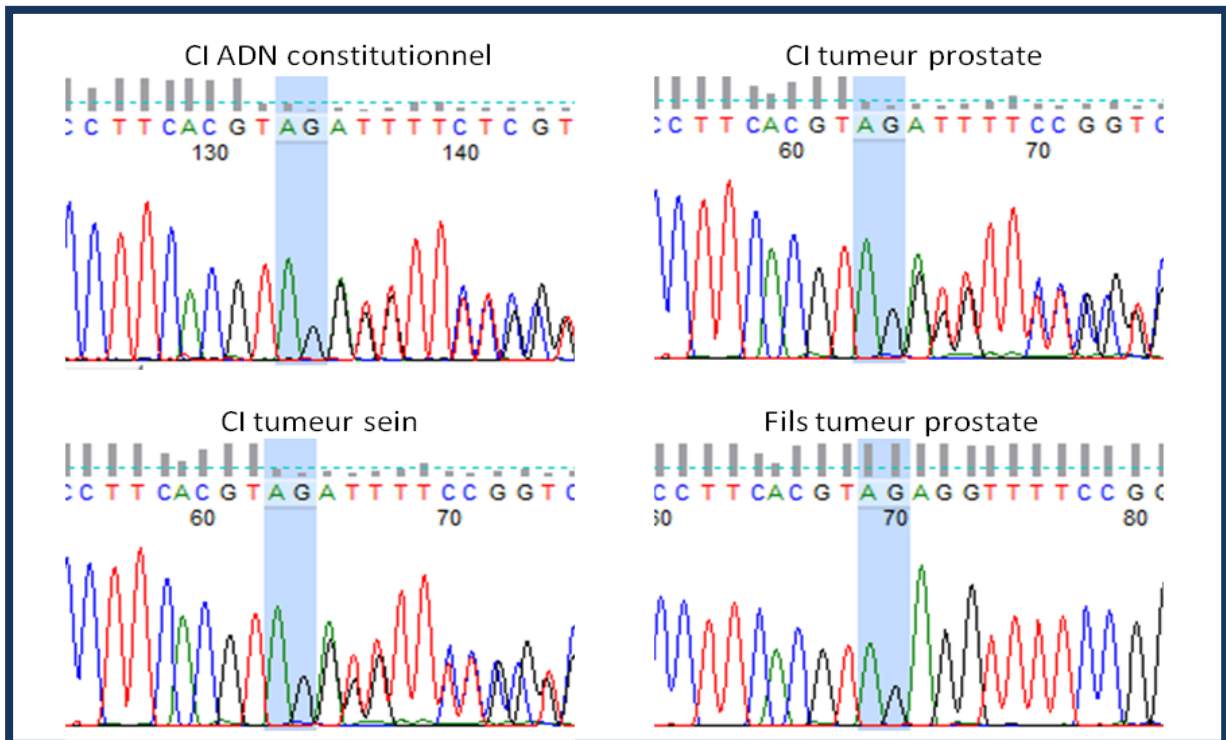


Figure 14: Séquençage Sanger de la variation de TP53I3 chez le cas index (CI) et son fils.

I.2.d. Expression protéique

L'analyse par Western blot de TP53I3 dans les fibroblastes du patient montre une réduction du taux de TP53I3 par rapport à 5 autres cultures primaires de fibroblastes (Figure 15).

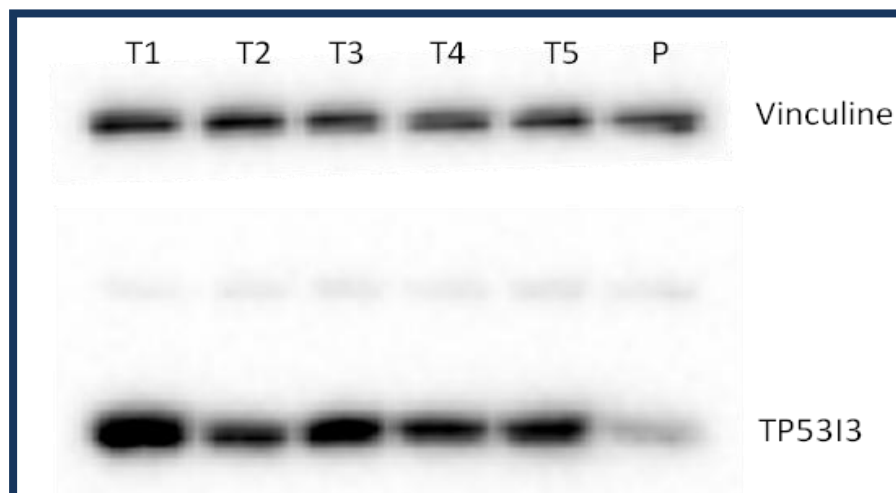


Figure 15: Analyse de l'expression protéique de TP53I3 par Western blot. T1 à T5: extraits protéiques de fibroblastes témoins, P: extrait protéique de fibroblastes du patient.

I.2.e. Effet de la variation sur la mort cellulaire

TP53/3 est impliqué dans l'apoptose induite par P53. Afin de vérifier si la variation de *TP53/3* et la diminution de l'expression qui l'accompagne ont un impact sur la mort cellulaire, nous avons cultivé les fibroblastes du patient et de 2 contrôles pendant 48h en présence de doses croissantes de doxorubicine une drogue génotoxique induisant l'apoptose. Nous avons ensuite évalué le taux de survie cellulaire par un test au XTT. Le taux de survie des cellules du patient ne montre pas de différence par rapport aux cellules contrôles (Figure 16). Elles ne sont donc pas résistantes à la mort cellulaire induite par un stress génotoxique.

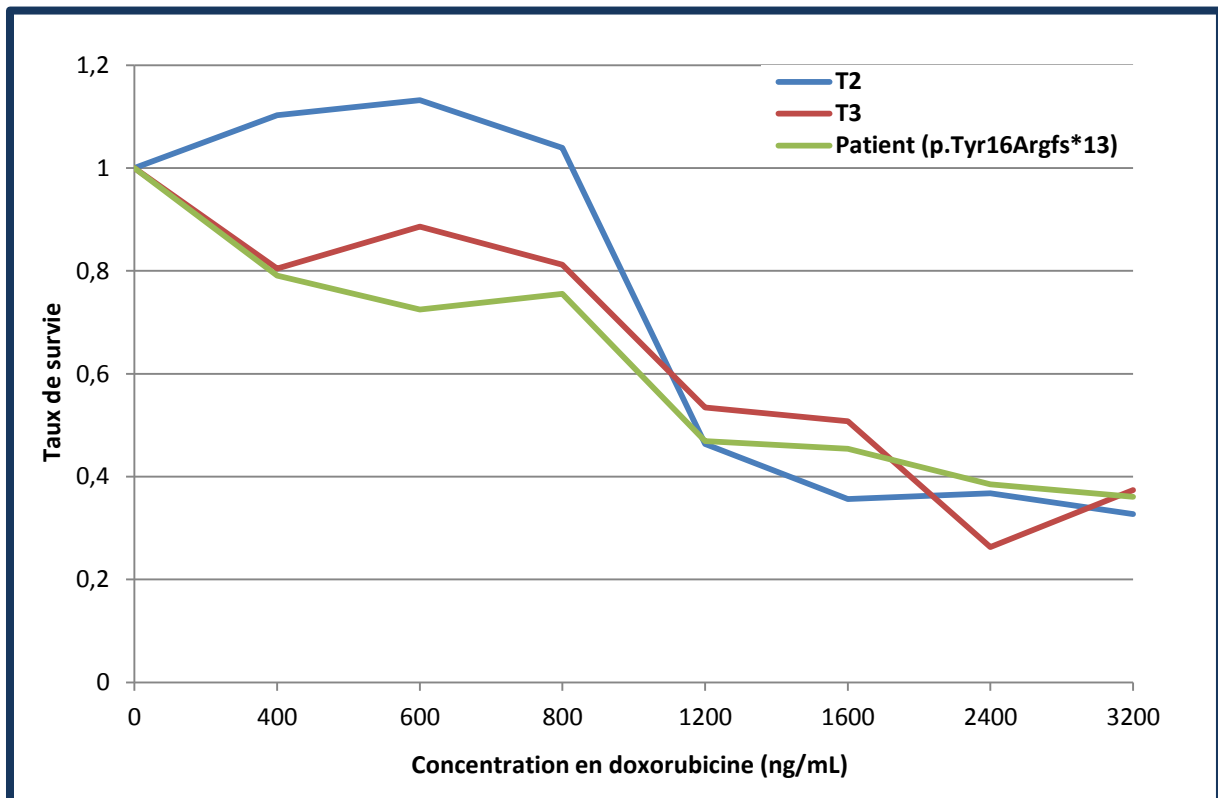


Figure 16: Effet de la variation de *TP53/3* sur la mort cellulaire induite par un stress génotoxique. Le taux de survie est calculé par rapport aux mêmes cellules cultivées dans les mêmes conditions dans un milieu sans doxorubicine. T2, T3: fibroblastes témoins sauvages

La variation de *TP53/3* entraîne une diminution de l'expression de la protéine mais n'a pas montré d'impact sur l'apoptose. Il semble donc que l'activité résultant de l'expression d'un seul allèle de *TP53/3* soit suffisante pour induire l'apoptose. Ces résultats peuvent également suggérer que *TP53/3* n'est pas indispensable pour déclencher l'apoptose et que des voies alternatives existent pour assurer cette fonction. De plus, l'expression de *TP53/3* est dépendante de P53 (Flatt et al., 2000). Or la Doxorubicine peut induire l'apoptose indépendamment de P53 ou avoir un mode d'action différent en fonction du type cellulaire (Tsang et al., 2003; Wang et al., 2004). Il pourrait donc être utile de répéter ce test avec d'autres molécules qui n'induisent

l'apoptose que par l'intermédiaire de l'activation P53 ou de tester l'effet d'une surexpression directe de P53 sur la survie des cellules du patient. *TP53I3* est également impliqué dans la réponse aux lésions de l'ADN. L'étude de Lee et al. montre notamment que la perte de *TP53I3* diminue la phosphorylation de CHK1 et CHK2, deux acteurs majeurs du contrôle du cycle cellulaire et de la réparation de l'ADN. Nous n'avons pas réalisé de tests sur cet aspect de la fonction de *TP53I3* mais il serait intéressant de vérifier si, comme pour ATR, la perte d'un allèle de *TP53I3* a un impact sur la phosphorylation de ces deux protéines. En attendant d'autres tests, la variation de *TP53I3* reste de signification inconnue.

I.3. DLC1 (Deleted in Liver Cancer 1)

Le troisième cas de CSM concerne un patient atteint d'un cancer du sein bilatéral à 54 et 56 ans. Son père a également eu un CSM à 75 ans. L'ES a révélé une variation faux-sens hétérozygote (c.2736C>G - p.His912Gln) du gène *DLC1*. Elle concerne un acide aminé bien conservé au cours de l'évolution et se trouvant dans une structure hélicale. Elle est absente des bases de données ExAC et gnomAD (Tableau 3).

DLC1 code pour une Rho-specific GTPase accelerating protein (RHO-GAP). Les RHO-GAP sont des inhibiteurs des RHO-GTPases elles mêmes impliquées dans différents processus cellulaires en lien avec la tumorigénèse. Leur dérégulation peut notamment perturber l'organisation des structure épithéliales, accroître la mobilité cellulaire, favoriser la dégradation de la matrice extra cellulaire et la progression dans le cycle cellulaire. Les mutations activatrices des RHO-GTPases sont rares mais elles sont souvent surexprimées dans les cellules cancéreuse (Sahai and Marshall, 2002). De ce fait, la perte de fonction des leur régulateurs comme les RHO-GAP peut être lié à la transformation tumorale. Une étude a montré que des variations faux-sens de *DLC1* engendrent une diminution de la stabilité de sa protéine et une augmentation de la viabilité cellulaire des cellules surexprimant les formes mutante de *DLC1* (Park et al., 2015).

Après la naissance, la glande mammaire se développe par morphogénèse ramifiée pour former un système canalaire, suivi d'une alvéologénèse pendant la grossesse. La structure clé à l'origine de ce processus est le bourgeon terminal où les précurseurs épithéliaux se développent et se différencient en cellules luminales et myoépithéliales formant les deux couches des canaux mammaires (Hinck and Silberstein, 2005; Silberstein, 2001). Comme dans tous les tissus épithéliaux, la polarisation des cellules est essentielle pour la formation des canaux et est influencée par des interactions avec la matrice extra cellulaire et les jonctions entre les cellules (Roignot et al., 2013). La perte de cette polarité cellulaire est une caractéristique des tumeurs épithéliales et peut jouer un rôle important dans leur initiation et leur progression (Wodarz and Näthke, 2007). Les Rho GTPases jouent un rôle essentiel dans la formation et le maintien des structures épithéliales et plusieurs études ont montré leur importance dans le développement des glandes mammaires (Chakravarty et al., 2003; Heckman-Stoddard et al., 2011; McCormack et al., 2013). La possible implication de *DLC1* dans une prédisposition au cancer du sein est fortement supportée par l'étude de Basak et collaborateurs montrant que des souris hétérozygotes pour *DLC1* présentent des anomalies morphologiques des glandes mammaires,

notamment une augmentation des ramifications canalaire et des déformations des bourgeons terminaux et des points de ramification. La culture tridimensionnelle des cellules épithéliales mammaires des souris hétérozygotes forme des canaux dont la lumière est obstruée. Ce phénotype est retrouvé quand *DLC1* est inhibé dans des cellules épithéliales mammaires normales. Cette étude montre également une augmentation de l'activité de la Rho-GTPase RhoA dans les cellules des souris hétérozygotes *DLC1 +/-*. Cela suggère que la réduction de l'expression de *DLC1* peut perturber la polarisation des cellules épithéliales et le développement des glandes mammaires chez la souris.

DLC1 est considéré comme un gène suppresseur de tumeur et son expression est diminuée dans différents cancers (Wang et al., 2016). Différents types d'altérations sont responsables de cette sous expression dans les cellules cancéreuses : au niveau génomique par des délétions de son locus, au niveau épigénétique par une hyper-méthylation de son promoteur ou au niveau protéique par l'augmentation de sa dégradation ou la perturbation de sa localisation (Durkin et al., 2007; Lukasik et al., 2011; Popescu and Goodison, 2014). Une étude fonctionnelle sur des cellules de cancer gastrique montre une augmentation de la viabilité cellulaire quand l'expression de *DLC1* est inhibé par un siRNA ou quand elles sur-expriment des mutants faux sens de *DLC1* (Park et al., 2015).

I.3.a. Séquençage ciblé sur la cohorte de réplication

Le séquençage de *DLC1* dans la cohorte de réplication a mis en évidence 4 variations, 3 faux-sens et 1 touchant un site d'épissage. Toutes les variations faux-sens sont connues dans Exac et gnomAD. Une seule (c.1448A>G) touche un domaine fonctionnel de la protéine mais ses scores de pathogénicité n'indiquent pas un effet délétère. La variation c.137A>G est prédite délétère par SIFT et a un score Polyphen élevé mais elle ne fait pas partie d'un domaine fonctionnel. La variation d'épissage (c.167+5G>C) est située sur le site donneur de l'exon 10 de *DLC1*. HSF prédit une forte probabilité d'impact sur l'épissage de cette variation. Parmi les 4 variations de la cohorte de réplication, nous avons décidé de ne retenir que cette dernière pour la suite de ce travail.

Tableau 6: Variations de *DLC1* identifiées dans la cohorte de réplication

Variation: génomique cDNA (NM_182643.1) protéine	Fréquences: ExAC gnomAD	Score de prédiction in silico - Prédiction d'impact sur l'épissage					Domaine - Structure
		CADD	polyPhen	GERP	Grantham	SIFT	
chr8:g.12955903C>G c.3167+5G>C -	Absent Absent	HSF: Altération du site donneur Très probablement affectant l'épissage					Polybasique cluster
chr8:g.13357444T>C c.137A>G p.Lys46Arg	3,62E-04 2,05E-04	CADD	polyPhen	GERP	Grantham	SIFT	-
		22.7	1	4.38	26	Délétère	
chr8:g.12968305A>G c.1448T>C p.Leu483Ser	5,77E-05 3,38E-05	CADD	polyPhen	GERP	Grantham	SIFT	Domaine SAM Structure hélicale
		12.20	0.027	-4.18	145	Toléré	
chr8:g.13357349T>C c.232A>G p.Met78Val	9,88E-05 7,08E-05	CADD	polyPhen	GERP	Grantham	SIFT	-
		21.4	0.587	3.79	21	Toléré	

I.3.b. Impact sur l'épissage

Deux variations du gène *DLC1* ont été identifiées, l'une par ES (p.His912Gln) et la seconde (c.3167+5G>C) dans la cohorte de réplication. Nous n'avons eu accès qu'à de l'ADN constitutionnel et de l'ARN pour les deux patients. La seconde variation touche le site donneur d'épissage de l'exon 10 et est prédite comme ayant très probablement un impact sur l'épissage. Pour vérifier cela, nous avons extrait l'ARN de la patiente et préparé deux amplicons couvrant respectivement les exons 9 à 11 et 9 à 12 du cDNA. L'électrophorèse sur gel d'agarose des produits de PCR ne montre pas de bandes pouvant correspondre à un transcrit altéré (Figure 17A). Le séquençage des produits de PCR montre une très faible proportion de transcrits avec rétention de 48 bases de l'intron 10 (0,5 % des transcrits pour l'amplicon couvrant les exons 9-11 et 0,7 % pour l'amplicon 9-12) (Figure 17B). Etant donné cette très faible proportion de transcrits altérés, il est normal qu'ils ne soient pas visibles sur gel et il est très peu probable que cette variation ait un effet significatif sur l'activité de *DLC1*.

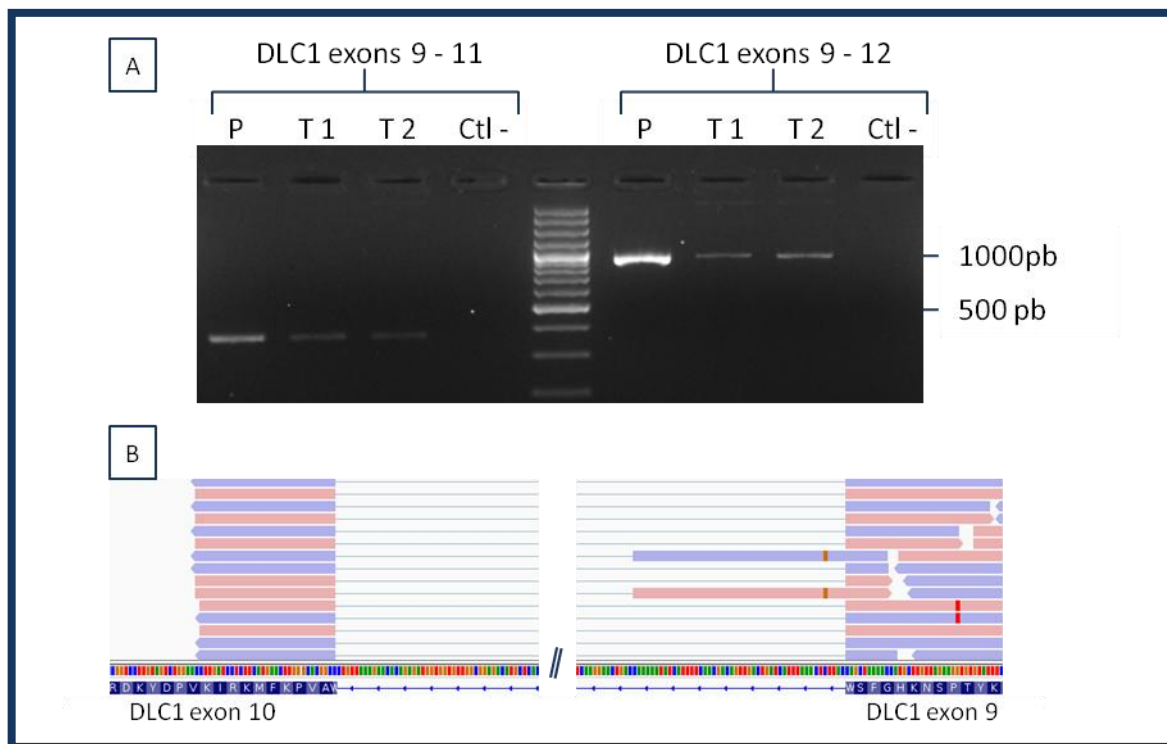


Figure 17: Etude de l'impact de la variation c.3167+5G>C sur l'épissage de *DLC1*. A: Migration sur gel d'agarose à 3% des produits de PCR. B: Visualisation de 2 lectures montrant la rétention de 48 bases de l'intron 9 de *DLC1*.

La variation d'épissage de *DLC1* peut être classée comme probablement bénigne. Toutes les autres variations de *DLC1* identifiées par ES ou dans la cohorte de réplication sont des faux-sens dont l'impact sur la fonction du gène est difficilement évaluable. Nous rechercherons activement d'autres variations perte de fonction de ce gène dans la cohorte de formes extrêmes de cancer du sein qui va s'enrichir avec la poursuite de ce projet.

II. Variations identifiées dans les autres présentations cliniques

II.1. Gènes connus comme étant responsables de prédisposition génétique au cancer

II.1.a. PALB2 (Partner and Localizer of BRCA2)

Une variation faux-sens hétérozygote du gène *PALB2* (p.Ile76Val) (Tableau 3) a été identifiée chez une patiente atteinte d'un cancer du poumon à 26 ans. Cette variation avait été classée comme de signification inconnue dans le compte rendu diagnostique d'un large panel de gènes. Son père a présenté un cancer de la prostate jeune (38 ans) et son grand père paternel a également développé un cancer du pancréas à 47 ans. L'exome à été réalisé en trio avec la tumeur. La variation est héritée du père.

PALB2, aussi connu sous le nom de *FANCN* (Fanconi anemia, complementation group N) joue un rôle majeur dans le processus de réparation de l'ADN par recombinaison homologue. Il a la capacité de recruter BRCA2 et RAD51 aux sites de lésions de l'ADN et permet l'assemblage du complexe BRCA1-PALB2-BRCA2 qui est essentiel pour la réparation par recombinaison homologue. Il est associé à une prédisposition au cancer du sein et du pancréas (Jones et al., 2009; Rahman et al., 2007). Cette variation a été rapportée une fois dans la littérature chez une patiente atteinte d'un cancer du sein (Foulkes et al., 2007) et 4 fois dans ClinVar. La variation est présente dans les bases de données ExAC et gnoAD à des fréquences relativement faibles (2.10^{-5} et 3.10^{-5} respectivement). L'acide aminé concerné se situe dans un domaine liaison à l'ADN et d'interaction avec BRCA1 et RAD51. Il est peu conservé au cours de l'évolution et la substitution d'une valine pour une isoleucine n'est pas considérée comme dommageable pour la fonction ou la structure de la protéine car ces deux acides aminés ont des propriétés similaires. Cependant, plusieurs modèles d'analyse de l'épissage prédisent que cette variation peut créer un site donneur cryptique dans l'exon 4 et altérer l'épissage normal de l'ARN messager. Cette variation reste de signification inconnue avant une éventuelle vérification de l'impact de la mutation sur l'épissage des ARN messagers de PALB2. L'organisation du recueil des prélèvements nécessaires est en cours.

II.1.b. BLM (BLM RecQ like helicase)

Le gène *BLM* code pour une hélicase impliquée dans les processus de réplication et de réparation de l'ADN. Il a également un rôle important dans la régulation des échanges de chromatides sœurs. Des mutations homozygotes ou hétérozygotes composites majoritairement tronquantes de *BLM* sont responsables du syndrome de Bloom (OMIM 210900), caractérisé par un retard de croissance pré et post natal, un érythème télangiectasique photosensible de la face, une susceptibilité accrue aux infections, et une instabilité génétique marquée par un taux élevé d'échange de chromatides sœurs. Les patients atteints de ce syndrome ont également une forte prédisposition à différents types de cancer. Les néoplasies les plus fréquentes sont des lymphomes et des leucémies durant l'enfance et l'adolescence et des adénocarcinomes du colon, de l'œsophage ou du sein à l'âge adulte.

Deux variations hétérozygotes du gène *BLM* ont été identifiées chez une patiente atteinte d'un adénocarcinome ovarien à 30 ans mais ne présentant pas d'autres caractéristiques du syndrome de Bloom. L'une est une variation faux sens substituant la proline en position 707 par une serine (p.Pro707Ser). Le second est une variation non sens en position 548 (p.Gln548*). Toutes deux sont connues dans ExAC et gnomAD à des fréquences respectives de 2.10^{-3} et 2.10^{-4} . L'exome a été réalisé en solo. Le père de la patiente a eu un cancer du pancréas à 74 ans et une de ses sœurs a eu un cancer du poumon à 51 ans. Tous deux étant décédés, il n'est pas possible de vérifier s'ils sont porteurs de ces variations.

Des mutations bi-alléliques du gène *BLM* sont responsables du syndrome de Bloom. Il est caractérisé par une instabilité chromosomique et une augmentation de risque de cancer. Afin de vérifier la phase de ces deux variations (conformation cis ou trans), elles ont été recherchées chez les deux filles de la patiente (Figure 18). L'une des deux est porteuse des deux variations et

la seconde n'en a aucune. Les deux variations sont donc en cis et ne peuvent pas conduire à une inactivation bi-allèle de *BLM*. Cela revient à ne considérer que la variation non-sens (p.Gln548*).

Deux études cas/contrôle ont montré une association entre cette variation et le développement de cancers du sein dans des population d'Europe de l'Est (Prokofyeva et al., 2013; Sokolenko et al., 2012). D'autres mutations tronquantes hétérozygotes de *BLM* ont été identifiées par ES dans des cancer du sein et du colon de phénotype extrême (Thompson et al., 2012; Voer et al., 2015). La fréquence de cette variation dans les bases de données de sujets sains ainsi que sa présence dans les groupes contrôles de différentes études démontrent une pénétrance faible. Ce gène n'est pas inclus dans les panels de gènes de prédisposition au cancer du sein, qui ne comprennent que des gènes à forte pénétrance, expliquant que ce résultat n'est pas été identifié dans le cadre du diagnostic. Il est possible que ce résultat ne puisse être qu'un facteur de prédisposition au sein d'une hérédité multifactorielle. Il est également possible que certaines variations du gène *BLM* aient une pénétrance plus élevée que d'autres. Les avancées scientifiques permettront sûrement à l'avenir d'améliorer les connaissances sur ce point.

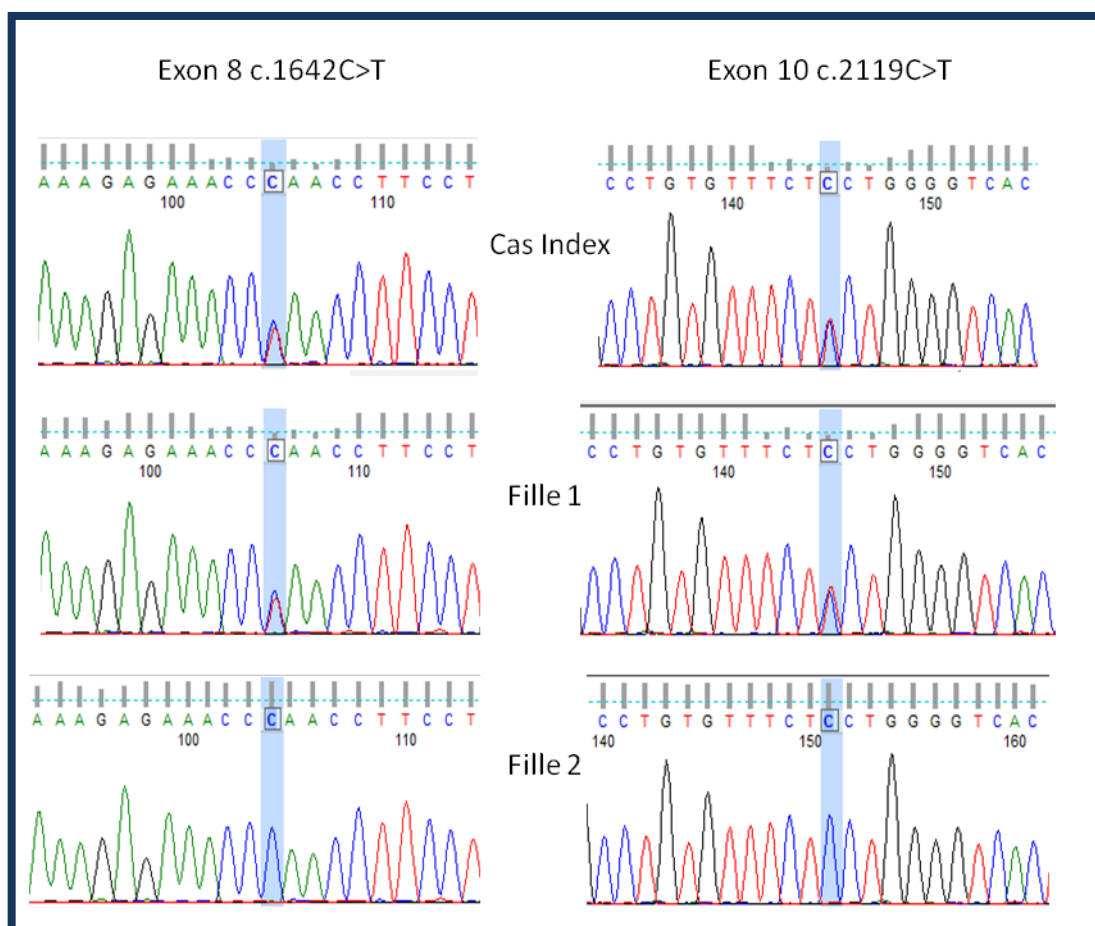


Figure 18: Confirmation des variations du gène *BLM* chez la patiente et ses deux filles par séquençage Sanger.

II.2. Nouveaux gènes candidats

II.2.a. BIK (BCL2-Intercating Killer)

Une délétion de 14 bases emportant les 9 dernières bases de l'intron 3 et les 5 premières bases de l'exon 4 a été identifiée dans le gène *BIK* (c.261-9_265delTGCTATAGCCTGG) (Tableau 3). Le patient a présenté des adénomes du colon à l'âge de 16 ans sans antécédents de cancer dans la famille. L'exome a été réalisé en solo car l'ADN des parents n'était pas disponible au moment de l'analyse. La ségrégation familiale a révélé que cette mutation était héritée de la mère. Le séquençage haut débit ciblé des exons 2 à 5 des ARN messagers de *BIK* montre l'apparition d'un transcrit alternatif avec une délétion des 28 premières bases de l'exon 4 dans environ 28 % des lectures (Figure 19). Cette délétion est hors phase et conduit à l'apparition d'un codon stop prématuré 23 acides aminés plus loin (p.Ser87Argfs*23).

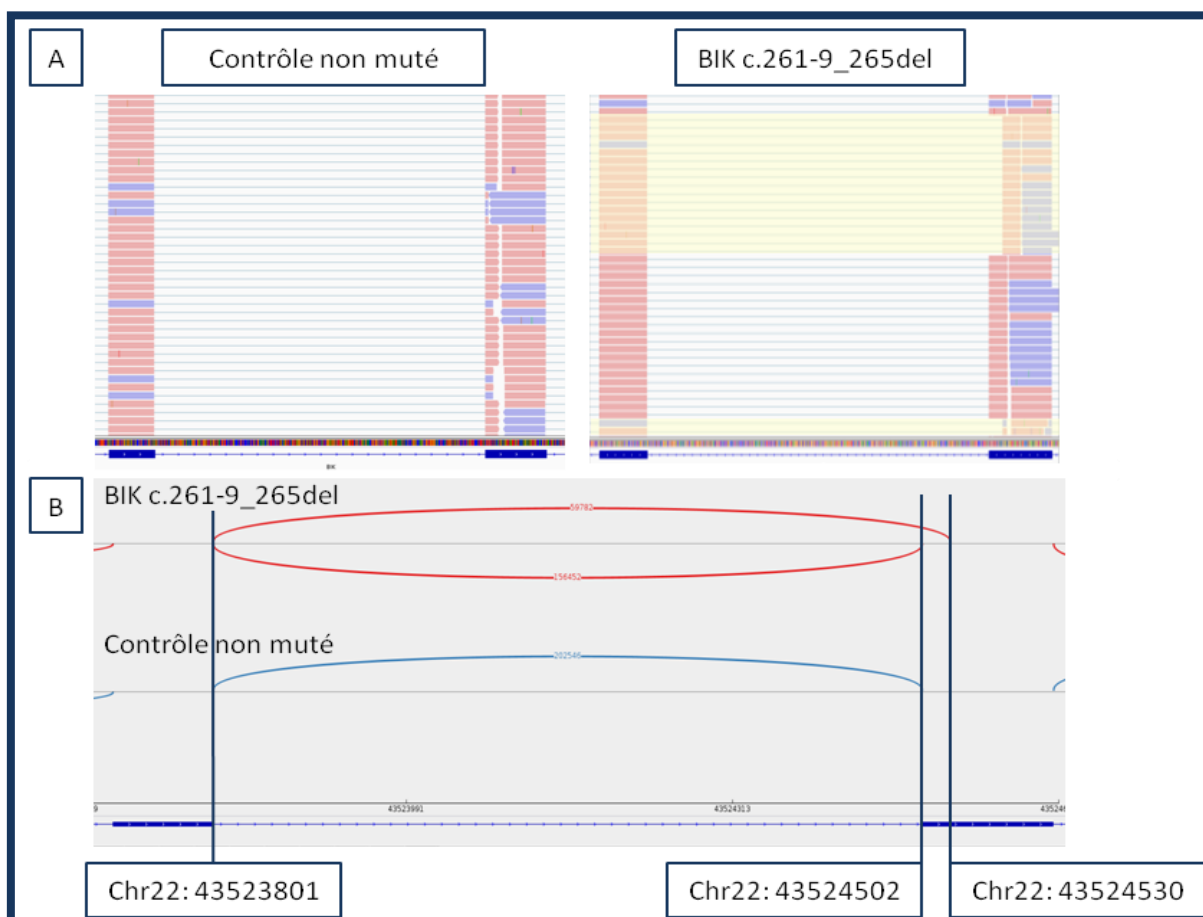


Figure 19: Séquençage ciblé de l'ARN messager de BIK. A: représentations IGV (integrated genome viewer) des lectures entre les exons 3 et 4. Les lectures dans les cadres jaunes montrent une perte de 28 premières bases de l'exon 4. B: Sashimi plot montrant la perte de 28 bases dans 28% des lectures.

BIK fait partie de la famille BCL2 qui regroupe des protéines pro et anti-apoptose. Elles ont toutes en commun un ou plusieurs domaines BH (BCL2 homology). Parmi les membres pro-apoptotiques, on trouve des effecteurs de l'apoptose comme BAX ou BAK, des activateurs comme BIM et BID et des "sensibiliseurs" dont fait partie BIK (Letai, 2017). Les activateurs et les sensibiliseurs ne possèdent qu'un unique domaine BH, le domaine BH3. Ils font partie d'une sous-famille appelée BH3-only. En réponse à différents signaux, les activateurs stimulent l'oligomérisation des effecteurs qui vont aller perméabiliser les membranes externes des mitochondries libérant ainsi le cytochrome c dans le cytoplasme pour déclencher la cascade d'apoptose. Les membres anti-apoptotiques comme BCL2 ou BCL-XL ont la capacité de séquestrer les activateurs ou les effecteurs activés pour inhiber l'apoptose. Les BH3-only "sensibiliseurs" peuvent se lier au domaine BH3 des membres anti-apoptose, ce qui prévient leur action inhibitrice (Certo et al., 2006; Cheng et al., 2001; Letai et al., 2002). Ils peuvent également libérer les activateurs ou les effecteurs déjà séquestrés par les inhibiteurs (Certo et al., 2006). Cette étape est considérée dans la plupart des cas comme le point de non-retour lors de l'engagement d'une cellule dans le processus d'apoptose (Letai, 2017).

Des études ont montré que l'expression de BIK est augmentée en cas de stress génotoxique. Cette expression peut être dépendante ou indépendante de la stabilisation de P53 (Mathai et al., 2002, 2005; Real et al., 2006). De plus, l'inhibition de BIK entraîne une diminution de l'apoptose dans la lignée de cancer colorectal HCT116 soumise à des drogues génotoxiques (Real et al., 2006). L'étude de l'expression de plusieurs membres de la famille BCL2 dans des cellules de cancer du sein dépendante des œstrogènes montre que seule l'expression de BIK est significativement affectée par la variation de la concentration en œstrogènes dans le milieu de culture et que BIK est nécessaire à la mort par apoptose provoquée par un traitement anti-œstrogène (Hur et al., 2004, 2006).

Différents types d'altérations somatiques de BIK ont été décrites: des délétions de son locus (del22q12-q13.2) dans des carcinomes rénaux, des cancers colorectaux, cancer des voies orales ou des gliomes (Bredel et al., 2005; Castells et al., 1999; Reis et al., 2002; Sturm et al., 2005), la méthylation de son promoteur dans des carcinomes rénaux et des lignées cellulaires de cancer du poumon et de la prostate (Dai et al., 2006; Sturm et al., 2005) ainsi que des mutations faux sens dans des lymphomes B (Arena et al., 2003).

Nous gardons en mémoire ce gène candidat, et nous resterons attentifs aux données de la littérature sur le sujet. En parallèle, nous augmentons significativement notre cohorte de formes extrêmes de cancers colo-rectaux séquencés en exome, au sein de laquelle nous pourrions rechercher bioinformatiquement des variations candidates du gène *BIK*.

II.2.b. DDFA (DNA Fragmentation Factor subunit Alpha)

Une délétion hétérozygote d'une base entraînant un décalage du cadre de lecture et l'apparition d'un codon stop prématuré (c.779delT - p.Leu260Trpfs*16) (Tableau 3) a été identifiée dans le gène *DDFA* chez une patiente ayant présenté un adénocarcinome de l'ovaire à 35 ans puis des polypes coliques et des nodules médiastinaux à 55 et 58 ans. Il n'y a pas

d'agrégation de cancer dans la famille. La patiente n'ayant pas de contact avec son père et sa mère étant décédée, il n'a pas été possible de réaliser l'exome en trio ni la ségrégation familiale de la variation.

Le gène *DFFA* code pour une protéine impliquée dans la fragmentation de l'ADN, une des caractéristiques de la mort cellulaire par apoptose. Toutefois, *DFFA* ne fragmente pas directement l'ADN. Elle forme, avec la protéine *DFFB*, le complexe *DFF* (DNA Fragmentation Factor) dans lequel elle joue le double rôle de chaperonne et d'inhibiteur de *DFFB*. Lors de la synthèse de *DFFB* dans les ribosomes, *DFFA* se lie à la chaîne peptidique et permet sa bonne conformation. Quand la synthèse est terminée, *DFFA* reste lié à *DFFB* et l'inhibe. Lorsqu'un stimulus apoptotique active les caspases, la caspase 3 vient cliver *DFFA* et libérer *DFFB* qui va alors pouvoir entrer dans le noyau et fragmenter l'ADN (Nagata, 2000). En l'absence de *DFFA*, aucune activité nucléase de *DFFB* n'est détectable, montrant le rôle essentiel de la fonction de chaperonne de *DFFA* (Zhang et al., 1998).

Une étude menée *in vitro* sur des cellules épithéliales coliques humaines et *in vivo* chez la souris montre que la déficience en *DFFA* diminue la sensibilité aux signaux de mort cellulaire et favorise l'instabilité génétique et l'apparition de tumeurs (Errami et al., 2013).

Différents types d'altérations somatiques de *DFFA* ont été décrites: une diminution de l'expression de l'ARN messager ou de la protéine dans des tumeurs du colon (Errami et al., 2013), de l'endomètre (Brustmann, 2007) et de l'œsophage (Konishi et al., 2002), des mutations faux-sens dans des neuroblastomes (Abel et al., 2002, 2004) ainsi que des délétions de son locus (1p36.2) dans différents types de cancer (neuroblastomes, syndromes myélodisplasiques, cancer du foie, sein, colon peau) (Bieche et al., 1998; Nakagawara et al., 2000; Schwab et al., 1996).

De même, nous gardons en mémoire ce gène candidat, et nous resterons attentifs aux données de la littérature sur le sujet. En parallèle, nous augmentons significativement notre cohorte de formes extrêmes de cancers ovariens séquencés en exome, au sein de laquelle nous pourrions rechercher bioinformatiquement des variations candidates du gène *DFFA*.

II.2.c. FAT4 (FAT atypical cadherin 4)

Deux variations faux-sens hétérozygotes (c.2203C>T - p.Arg735Trp et c.9997G>A - p.Val3333Ile) ont été identifiées dans le gène *FAT4* chez une patiente présentant un cancer du sein bifocal à 30 ans. Deux de ses tantes ont également eu un cancer du sein précoce à 30 et 40 ans. Chaque variation n'est rapportée qu'une seule fois dans les bases de données ExAC et gnomAD (Tableau 3).

FAT4 code pour une protéine transmembranaire de la famille des Cadherin-related proteins. Elle contient 34 domaines cadhérines extracellulaires. Cependant, contrairement aux cadhérines classiques (E-cadherin, N-cadherin...), elle n'a pas une fonction d'adhésion cellulaire mais joue un rôle de molécule de signalisation. Elle est impliquée dans la régulation de la prolifération cellulaire, le contrôle de la taille de organes et la polarité des cellules via la voie de signalisation Hippo (Thomas and Strutt, 2012).

Plusieurs études ont montré que *FAT4* est un gène suppresseur de tumeur notamment dans des cancers du sein et des cancers gastriques. La perte d'expression de *FAT4* dans les cellules épithéliales mammaires non cancéreuses NOG8 leur confère un pouvoir cancérogène et la réexpression de *FAT4* dans les cellules issues de ces tumeurs inhibe fortement leur croissance (Qi et al., 2009). L'inhibition de l'expression de *FAT4* favorise la prolifération et les capacités de migration et d'invasion de plusieurs lignées cellulaires de cancer du sein triple négatif (Hou et al., 2016). Ces différents processus sont affectés de la même façon dans des lignées cellulaires de cancer gastriques via l'activation de la voie Wnt/ β -catenin (Cai et al., 2015). Deux de ces études montrent également que la perte d'expression de *FAT4* inhibe l'expression de marqueurs épithéliaux et favorise l'expression de marqueurs mésenchymateux (Cai et al., 2015; Hou et al., 2016). Cela suggère que *FAT4* régule la progression des cancers en inhibant la transition épithélio-mésenchymateuse.

Au niveau somatique, l'expression de l'ARN messager de *FAT4* est souvent diminuée par méthylation de son promoteur dans les cancers du sein et du poumon (Qi et al., 2009; Rauch et al., 2012) et des variations ou délétions de *FAT4* sont fréquemment détectées dans les mélanomes, les cancers du pancréas, les carcinomes épidermoïdes de la tête et du cou et les cancers gastriques (Katoh, 2012).

De même, nous gardons en mémoire ce gène candidat, et nous resterons attentifs aux données de la littérature sur le sujet. En parallèle, nous augmentons significativement notre cohorte de formes extrêmes de cancers du sein séquencés en exome, au sein de laquelle nous pourrions rechercher bioinformatiquement des variations candidates du gène *FAT4*.

II.2.d. SETD2 (SET domain containing 2, histone lysine methyltransferase)

Une variation faux sens *de novo* du gène *SETD2* (c.4973C>T - p.Ser1658Leu) a été identifiée chez une patiente ayant présenté différentes néoplasies à 10 ans d'intervalle depuis l'âge de 7 ans: un ostéoblastome sacré à 7 ans, une tumeur osseuse bénigne à 17 ans, un pseudomyxome péritonéal à 27 ans et un macroadénome hypophysaire à 37 ans (Tableau 3). La patiente présente également des paumes et des plantes très sèches, une dent surnuméraire, une lordose marquée, un périmètre crânien à la limite supérieure de la normale et un comportement particulier. En l'absence d'antécédents de cancers au niveau familial, la stratégie d'exome en trio a été adoptée pour ce cas.

SETD2 est le seul gène humain connu capable d'apposer une marque de triméthylation sur la lysine 36 de l'histone H3 (H3K36me3) (Edmunds et al., 2008; Strahl et al., 2002). La perte de *SETD2* entraîne une absence complète de cette marque de méthylation sans perturber les niveaux de mono ou diméthylation de cet acide aminé. Sa fonction est donc très spécifique. La marque H3K36me3 peut ensuite être reconnue par des protéines effectrices appelées "*readers*", qui déterminent la conséquence fonctionnelle des modifications d'histones en recrutant des complexes protéiques impliqués dans des processus spécifiques comme la transcription, l'épissage des ARN messagers et la réparation de l'ADN (Li et al., 2016a). Plus récemment, il a été

démonstré que SETD2 à aussi une fonction de méthylation des microtubules. Il appose une marque de triméthylation sur la lysine 40 de l' α -tubuline des microtubules (α TubK40me3). La perte de cette marque provoque des défauts du fuseau mitotique et de la cytotérière ainsi que la formation de micro-noyaux et l'apparition de polyploïdies (Park et al., 2016). SETD2 peut également interagir avec le domaine N-terminal de P53 et moduler l'expression de plusieurs gènes cible de P53 incluant notamment les gènes d'apoptose *BCC3* (puma), *PMAIP1* (noxa) et *TP53AIP1* (Xie et al.).

SETD2 est localisé sur le bras court du chromosome 3 (3p21.31). La délétion de ce locus est un événement précoce et très fréquent dans les cancers du rein à cellules claires (ccRCC: clear cell Renal Cell Carcinoma) ((Kovacs et al., 1987). Des mutation de *SETD2* sont également retrouvées dans les cancers du poumon (Walter et al., 2017), de la prostate (Cerami et al., 2012), des gliomes (Fontebasso et al., 2013) et des leucémies (Zhu et al., 2014). L'études des gènes de ce locus à permis d'identifier plusieurs gènes suppresseurs de tumeur dont *SETD2* (Duns et al., 2010). La perte de fonction de SETD2 perturbe la structure des nucléosomes, la distribution des réplisomes , la progression des fourches de réplication, et la réparation de l'ADN par recombinaison homologue ou par le système MMR (Kanu et al., 2015; Li et al., 2013). La dérégulation de toutes ces fonctions cellulaires importantes peut contribuer à l'instabilité génomique, caractéristique des cellules tumorales. Enfin, l'inactivation de SETD2 dans les cellules épithéliales tubulaires primaires rénales permet de passer outre la barrière de la sénescence, facilitant ainsi la transformation maligne vers le ccRCC (Li et al., 2016b).

La variation identifiée dans *SETD2* est absente des bases de données ExAC et gnomAD. Elle touche un acide aminé bien conservé au cours de l'évolution. Elle se situe dans le domaine SET, responsable de l'activité méthyl-transférase de SETD2. Ce domaine est indispensable pour l'apposition des marques H3K36me3 et α TubK40me3 (Park et al., 2016). La base de données COSMIC montre un enrichissement en variations faux sens dans ce domaine. La même région est relativement dépourvue de variations dans la population saine indiquant une pathogénicité accrue des variations faux sens de ce domaine (Li et al., 2016a).

A noter que le gène SETD2 est également responsable du syndrome de Luscan-Lumish (OMIM 616831), associant macrocéphalie et troubles neurocognitifs. Compte tenu de la présentation clinique de la patiente, du caractère de novo de cette variation et de ces données bioinformatiques, ce gène nous semble être un bon gène candidat. Il nous semble pertinent de réaliser des études fonctionnelles, pour laquelle la patiente a accepté des prélèvements dans le cadre de la recherche. Nous cherchons actuellement à identifier une équipe avec qui une collaboration pourrait être mise en place pour aller plus loin.

Discussion

I. Cancers du sein masculin

L'ES et le séquençage ciblé sur la cohorte de réplication nous ont permis d'identifier des variations perte de fonction chez 5 hommes atteints d'un cancer du sein, 2 dans le gène *ATR*, 2 dans *DLC1* et 1 dans le gène *TP53/3*. L'inclusion d'*ATR* dans le panel de GPC du CGFL a révélé une troisième variation d'*ATR* dans un cancer du sein féminin. Des analyses supplémentaires ont été réalisées afin d'estimer l'impact des variations sur la fonction des gènes et leur possible implication dans une prédisposition génétique au cancer du sein. Le travail le plus abouti concerne le gène *ATR* pour lequel nous avons réalisé des analyses fonctionnelles en collaboration avec l'équipe du Pr Mark O'Driscoll de l'université du Sussex, spécialisée dans l'étude des systèmes de réparation de l'ADN.

En raison de sa fonction et de sa structure relativement proche de celle du gène *ATM*, déjà impliqué dans une prédisposition au cancer du sein, *ATR* a été considéré comme un gène candidat potentiellement responsable d'une prédisposition au cancer du sein. Avant ce travail, il a été testé sur deux cohortes, l'une finlandaise de 126 familles avec un risque de cancer du sein ou de l'ovaire estimé de modéré à fort (Heikkinen et al., 2005), l'autre canadienne comprenant 54 patients issus de familles à haut risque de cancer du sein et de l'ovaire négatifs pour *BRCA1/2* (Durocher et al., 2006). Aucune variation délétère n'a été trouvée dans chacune de ces études. Plus récemment, une équipe australienne a publié les résultats du séquençage d'un panel de gènes incluant *ATR* sur 2000 patients atteints de cancer du sein avec une forte histoire familiale, négatifs pour *BRCA1/2*, et 1997 contrôles non atteints. Ils ont identifié 3 variations perte de fonction chez les patients et une dans la population contrôle (Thompson et al., 2016). Enfin, une étude se rapprochant de notre projet a identifié une variation d'*ATR* dans l'exome d'un patient négatif pour un panel de gène de prédisposition au cancer du sein. Deux variations additionnelles ont été mises en évidence dans une cohorte de réplication comprenant 1717 patients (Spurrell, 2013). Le fait qu'aucune variation n'ait été trouvée dans les deux premières études peut être expliqué par la petite taille des cohortes et une sélection moins stricte des patients. Au total, en prenant en compte les deux dernières études citées et la nôtre, 8 variations tronquantes d'*ATR* ont été identifiées sur 3880 patients sélectionnés pour une prédisposition au cancer du sein, ce qui peut être en faveur d'une prédisposition génétique au cancer du sein ultra-rare. Aucune étude fonctionnelle de l'impact de ces variations n'a été réalisée dans les autres études.

Nous avons eu accès à des cellules de deux des patients mutés pour *ATR*. Les tests fonctionnels montrent une diminution de l'expression protéique d'*ATR* et une diminution de la phosphorylation de *CHK1*. *CHK1* est activé par *ATR* par phosphorylation des sérines 317 et 345. Il permet ensuite de diffuser le signal de l'activation d'*ATR* dans la cellule pour aboutir à la régulation de la réplication de l'ADN en contrôlant les fourches de réplication et la transcription de gènes ou l'activation de nombreux facteurs impliqués dans le contrôle du cycle cellulaire (Iyer

and Rhind, 2017). Même en l'absence de stress génotoxique, CHK1 est activé à un niveau basal et régule la réplication de l'ADN durant la phase S du cycle cellulaire. La perte de cette fonction nuit à la stabilité du génome. En cas de lésion dans l'ADN, un niveau d'activation plus important de CHK1 est nécessaire pour arrêter le cycle cellulaire et permettre la réparation (Iyer and Rhind, 2017). La diminution de la phosphorylation de CHK1 que nous avons mis en évidence dans les cellules mutées pour *ATR* peut donc avoir un impact significatif le contrôle du cycle cellulaire et la réparation de l'ADN.

L'ACPA et le séquençage ciblé des variations dans l'ADN tumoral n'a pas mis en évidence de perte d'hétérozygotie dans la tumeur du patient 1. Ces données sont concordantes avec la littérature suggérant qu'*ATR* est indispensable à la survie cellulaire et que son inactivation bi-allélique entraîne une instabilité génomique trop importante qui conduit à la mort cellulaire (Brown and Baltimore, 2000; Cortez et al., 2001; de Klein et al., 2000). En revanche, une augmentation de l'incidence des tumeurs a été observée chez des souris hétérozygotes pour *ATR* en comparaison avec leurs congénères non mutées (Brown and Baltimore, 2000).

Des mutations constitutionnelles homozygotes ou hétérozygotes composites d'*ATR* sont responsables d'un syndrome de Seckel, caractérisé par un retard de croissance intra-utérin, un nanisme, une microcéphalie avec déficience intellectuelle et des particularités faciales dites en "tête d'oiseau". Toutes les mutations décrites sont responsables de défaut d'épissage et d'une forte diminution de l'expression d'*ATR* mais, en accord avec les études in vitro ou sur un modèle murin, l'expression d'*ATR* n'est jamais nulle. Les cellules des patients montrent des défauts de la phosphorylation des substrats d'*ATR* induite par des dommages dans l'ADN, une instabilité des fourches de réplication et une perturbation de l'activation du point de contrôle G2/M du cycle cellulaire (Mokrani-Benhelli et al., 2013; O'Driscoll et al., 2003; Ogi et al., 2012).

Une mutation faux sens d'*ATR* a été identifiée chez plusieurs membres d'une famille avec un syndrome associant des télangiectasies cutanées, des anomalies des cheveux, des sourcils, des dents et des ongles (Tanaka et al., 2012). Dix des 24 membres de la famille atteints ont développé un cancer oropharyngé à partir de 30 ans pour le plus précoce. D'autres formes cancer sont présentes dans la famille incluant trois cancers de la peau, un cancer du sein à 35 ans et un cancer de l'utérus à 28 ans. Cette variation n'entraîne pas de diminution de l'expression d'*ATR* mais le séquençage de la tumeur d'un des individus atteints montre une perte de l'allèle sauvage d'*ATR*.

Nos résultats et les données de la littérature convergent vers un modèle dans lequel un stress répliatif modéré dû à l'haploinsuffisance d'*ATR* permet la survie des cellules mais engendre une instabilité génétique latente favorisant l'apparition de mutations somatiques responsables de la transformation tumorales des cellules. Cela concorde avec un des modèles d'haploinsuffisance obligatoire énoncé dans l'article de Berger et collaborateurs, selon lequel la perte des deux allèles d'un gène suppresseur de tumeur conduit à la mort cellulaire ou la sénescence plutôt qu'à la transformation tumorale (Berger et al., 2011). Ce mécanisme de transformation tumorale semble être différent de celui décrit par Tanaka et collaborateurs dans la famille avec prédisposition au cancer oropharyngé puisque la mutation faux-sens identifiée

n'entraîne pas de diminution de l'expression et qu'elle est associée à une perte d'hétérozygotie dans la tumeur (Tanaka et al., 2012).

Les données moléculaires et fonctionnelles collectées pour le patient 1 permettent de considérer la variation d'*ATR* comme la cause probable de son cancer. La réplication de ces analyses sont en cours pour le patient 3 et devraient confirmer l'implication d'*ATR* dans la prédisposition au cancer.

Etant donné que la variation du patient 2 ne conduit qu'à la perte de deux acides aminés et qu'il n'a pas été possible de réaliser d'analyses fonctionnelles, elle reste de signification inconnue.

Pour les deux autres gènes identifiés dans les CSM (*DLC1* et *TP53/3*), les analyses complémentaires ont été réalisées localement. Leurs résultats sont plus préliminaires et ne permettent pas de conclure sur l'implication du gène dans une prédisposition au cancer.

II. Gènes connus comme étant responsables de prédisposition génétique au cancer

PALB2 est un gène de prédisposition au cancer du sein bien établi. Les études estimant le risque de CS chez les femmes porteuses de *PALB2* rapportent un risque très élevé voisin de celui observé chez les femmes porteuses de mutations des gènes *BRCA1* ou *BRCA2* (Antoniou et al., 2014; Couch et al., 2017; Thompson et al., 2016). La variation identifiée dans *PALB2* concerne une patiente atteinte d'un cancer du poumon. Il n'existe pas à ce jour d'estimation du risque de cancer du poumon lié à des mutations de ce gène. Il s'agit d'une variation faux-sens ayant un potentiel impact sur l'épissage. Nous essayons d'obtenir de nouveaux prélèvements afin de vérifier cette prédiction sur l'ARN de la patiente. Cela permettrait de mieux caractériser l'effet d'une variation déjà décrite mais restant de signification inconnue et d'ouvrir une voie vers l'élargissement du spectre des cancers liés à une mutation de *PALB2*. Nous restons attentifs en parallèle à l'éventuelle reclassification de cette variation à l'avenir dans les bases de données, qui serait issue de travaux collaboratifs internationaux.

Des mutations délétères hétérozygotes de *BLM* ont été associées à une augmentation du risque de cancer du sein (Prokofyeva et al., 2013; Sokolenko et al., 2012; Thompson et al., 2012) et de CCR (Voer et al., 2015). En revanche, cette association est moins claire pour le cancer de l'ovaire ou de la prostate (Antczak et al., 2013; Bogdanova et al., 2015). Une étude *in vivo* sur un modèle murin hétérozygote pour une mutation délétère du gène homologue de *BLM* montre que les souris hétérozygotes développent des lymphomes plus tôt que leurs congénères sauvages après injection du virus de la leucémie murine, et développent deux fois plus de tumeurs intestinales quand elles sont croisées avec une lignée de souris porteuse d'une mutation du gène homologue du suppresseur de tumeur *APC*. L'absence de perte d'hétérozygotie au locus de *BLM* dans ces tumeurs montre que l'haploinsuffisance de *BLM* est suffisante pour induire ces effets (Goss et al., 2002). En accord avec cette observation, un étude a démontré que les tumeurs du sein de patientes hétérozygotes pour *BLM* ne montrent pas

d'inactivation du second allèle (Suspitsin et al., 2014). Les données de la littérature semblent indiquer que l'haploinsuffisance de BLM peut au moins être un facteur modificateur favorisant le développement de cancers.

Dans notre approche, nous nous attendions à retrouver l'implication de gènes connus comme étant responsables de prédisposition génétique de certains cancers dans d'autres types de cancer. En effet, le spectre de cancer décrit en lien avec un gène de prédisposition comprend d'abord les cancers à forte pénétrance, puis il est nécessaire d'avoir des études beaucoup plus importantes pour identifier des risques de cancers moins élevés. Nous avons appris cela de l'expérience du gène *BRCA2*, initialement décrit comme étant seulement responsable d'un excès de risque de cancer du sein et de l'ovaire, puis du cancer de la prostate (van Asperen et al., 2005; Edwards et al., 2003), puis du cancer du pancréas (van Asperen et al., 2005; Murphy et al., 2002). Un excès de risque de mélanome a été plus récemment discuté (Gumaste et al., 2015), et un sur-risque de CCR reste débattu (Oh et al., 2018; Pearlman et al., 2017; Sopik et al., 2015; Yurgelun et al., 2017). Ces premiers résultats nous poussent à être attentifs sur cette hypothèse.

Une 2^{ème} hypothèse était celle de l'implication de variations hétérozygotes de gènes responsables de prédispositions génétiques prouvées à l'état homozygote ou hétérozygote composite, comme cela a été identifié pour le gène *ATM* (Renwick et al., 2006). Dans ce cas, les conclusions sont plus difficiles car il ne peut s'agir que d'un facteur de prédisposition au sein d'une hérédité multifactorielle, bien que l'existence de certaines variations à plus forte pénétrance que d'autres reste possible, comme cela est le cas pour le gène *CHEK2* (Meijers-Heijboer et al., 2002). Les avancées scientifiques permettront sûrement à l'avenir d'améliorer les connaissances sur ce point.

III. Gènes candidats et résultats négatifs

Les variations identifiées dans les autres gènes candidats sont de signification inconnue et des analyses fonctionnelles ou la recherche de répliqués sont nécessaires pour déterminer si ces gènes peuvent être impliqués dans une prédisposition au cancer. Par exemple, nous avons accès à du tissu FFPE pour la patiente porteuse de la variation du gène *SETD2*. Nous allons commencer par vérifier, par immunohistochimie, si la variation entraîne une diminution de la modification H3K36me3 spécifique de SETD2.

Cette approche de phénotype extrême de cancer va être poursuivie dans les années à venir, et même étendue à un recrutement à l'ensemble de l'interrégion, ce qui permettra progressivement d'augmenter significativement la cohorte de formes extrêmes de cancers séquencés en exome, au sein de laquelle nous pourrons rechercher bioinformatiquement des variations candidates dans nos gènes d'intérêt. En parallèle, une veille de la littérature sera faite pour l'ensemble de ces gènes.

Une variation candidate a été identifiée dans un tiers des cas (9/27). La plupart des cas restent donc inexplicables après séquençage d'exome. Plusieurs hypothèses peuvent l'expliquer :

- Tout d'abord, bien que les cas sélectionnés présentent des phénotypes extrêmes, on ne peut pas exclure qu'une proportion de ces cancers ne soient pas dus à une prédisposition génétique à forte pénétrance mais soit le résultat d'une accumulation stochastique de variations somatiques. De nombreux facteurs environnementaux pouvant entrer en jeu et favoriser ce processus (habitudes alimentaires, exposition à des toxiques...). Une autre hypothèse est que certains cancers suivent le modèle d'une accumulation de variations communes à faible pénétrance telles que celles identifiées par les GWAS. Ce type de variation n'entre pas dans le spectre de ce projet. Elles se situent souvent en dehors des régions codantes, donc non capturées par l'ES, ou sont éliminées par les filtres de fréquences appliqués lors du variant calling.

- Il peut être compliqué d'interpréter certains résultats en raison d'un manque de connaissances scientifiques. Par exemple, dans un exome trio, une variation tronquante *de-novo* a été identifiée dans le gène *SLAIN1* mais il n'y a que très peu d'informations concernant la fonction de ce gène dans la littérature. Il est donc compliqué d'évaluer s'il pourrait être impliqué dans une prédisposition au cancer. Pour palier à ces difficultés, les données d'ES peuvent être ré-analysées périodiquement avec des outils bio-informatiques mis à jour et intégrant les dernières versions des bases de données. Cela peut permettre d'identifier de nouvelles variations ou de mieux interpréter des variations de significations inconnues à la lumière des publications les plus récentes. L'intérêt de cette ré-analyse a été démontré dans le domaine des maladies rares (Nambot et al., 2018).

- La cause de l'échec peut également être d'ordre technique. En effet, l'étape de capture des régions codantes du génome n'est pas parfaite en particulier dans des régions riches en GC. Le taux de couverture moyen des bases codantes à une profondeur minimale de 10X de notre cohorte est de 95%. Il y a donc 5% des régions codantes pour lesquels la couverture est insuffisante pour détecter efficacement les variations.

- Enfin, la prédisposition génétique pourrait être due à des variations non codantes, qui ne sont pas capturées par l'ES. Les variations non codantes peuvent avoir plusieurs effets. Des SNV, des duplications ou des délétions peuvent conduire à la création ou la perte de motifs de liaison de facteurs de transcription dans les régions promotrices. Des variants structuraux peuvent juxtaposer anormalement un élément de régulation et un oncogène ou un gène suppresseur de tumeur. Des variations peuvent également toucher des microARN ou leurs sites de liaison et conduire à une surexpression de leur gènes cibles. La perte de pseudogènes peut augmenter l'inhibition de l'expression du gène normal (Khurana et al., 2016). Des mutations de prédisposition au cancer, constitutionnelles, rares et non codantes ont déjà été décrites. On peut citer en exemple les mutations du promoteur du gène *TERT* associées à une forme de mélanome héréditaire (Horn et al., 2013) ou une mutation intronique du gène *BRCA2* responsable d'un épissage aberrant (Bakker et al., 2014).

Le séquençage de génomes complets (WGS: Whole Genome Sequencing) pourrait permettre de détecter ce type de variations. La forte diminution des coûts de séquençage et l'amélioration des connaissances concernant l'ADN non codant notamment grâce au projet ENCODE (ENCyclopedia Of DNA Elements) (Davis et al., 2018) rendent l'utilisation de cette technique envisageable. Le WGS peut être couplé au RNASeq (séquençage du transcriptome). Cela permet de mettre en évidence les conséquences des variations non codantes identifiées en WGS sur l'expression ou la transcription des gènes et facilite leur interprétation. De plus, le WGS ne nécessitant pas d'étape de capture, il offre donc une meilleure couverture de régions difficiles et est plus efficace pour la détection de grandes variations chromosomiques. Il est tout à fait possible que ce projet évolue vers le séquençage de génomes. Un financement spécifique a été demandé pour le séquençage de génome en trio dans les formes extrêmes de CCR afin d'évaluer les performances de cette approche.

Conclusion générale et perspectives

Le séquençage d'exome de 27 cas de cancer de phénotypes extrêmes a permis d'identifier 9 pistes de gènes candidats pour une prédisposition génétique au cancer. Ces premiers résultats montrent la pertinence d'utiliser l'ES et une approche de phénotypes extrêmes pour identifier de nouveaux gènes de prédisposition au cancer. Cette approche devra être poursuivie dans le temps dans le cadre de l'étude EXTRICAN, et qui sera ouverte en interrégional à partir du mois de Juin 2019. Quinze cas supplémentaires ont déjà été identifiés lors des dernières RCP. L'ES est en cours. Les données seront bientôt disponibles.

Le principal résultat de ce travail concerne le gène ATR. Son rôle central dans les processus de réparation de l'ADN et de contrôle du cycle cellulaire en fait un bon gène candidat pour une prédisposition au cancer du sein. Les analyses fonctionnelles qui ont pu être réalisées à partir de lignées cellulaires de deux patients mutés ont permis de valider l'impact des variations identifiées sur sa fonction et de classer ces variations comme probablement pathogènes. La pénétrance associée est totalement inconnue, et compte tenu de sa faible fréquence potentielle, nécessitera des études larges internationales.

Les variations identifiées dans les autres présentations cliniques sont de signification inconnue et nécessitent d'autres arguments pour valider l'hypothèse de leur implication dans une prédisposition génétique au cancer. Les perspectives suivantes sont envisagées :

- D'autres analyses fonctionnelles sont envisagées pour certains gènes candidats.
- La constitution de cohortes de réplication pour rechercher des récurrences pour les autres présentations cliniques que les CSM est à discuter, puisque l'un des arguments les plus forts en faveur de la pathogénicité d'une variation est l'identification d'autres variations du même gène chez d'autres patients présentant un phénotype similaire.
- La continuité de ce projet va à terme permettre d'avoir un nombre plus important de patients avec phénotype extrême, et donc d'éventuelles répliquions.
- Le partage anonyme des données cliniques et moléculaires sur des plateformes de soumission en ligne est une autre méthode utilisée dans l'étude des maladies rares pour rechercher des répliquions. Les variations de signification inconnues identifiées dans ce projet ont été déposées sur la plateforme GeneMatcher (Sobreira et al., 2015). Cette pratique n'étant pas courante dans le domaine de l'oncogénétique, aucune répliquion n'a encore été identifiée par ce biais mais cela pourrait arriver dans le futur si les méthodes issues de l'étude des maladies génétiques rares prennent de l'importance en oncogénétique et conduisent à l'identification de variations rares.

Sur le plan personnel, ce projet m'a permis d'approfondir mes connaissances dans le domaine de l'oncogénétique, notamment sur la prise en charge des patients chez qui une

prédisposition au cancer est suspectée et l'architecture génétique de la prédisposition au cancer. L'interprétation de données d'exome m'a fait comprendre les difficultés liées à l'interprétation des données de séquençage à grande échelle et la nécessité d'utiliser des bases de données et des outils bioinformatiques d'annotation et de prédiction de la pathogénicité des variations pour les classer et les prioriser. Cela m'a également permis de comprendre l'intérêt des différentes stratégies de séquençage pour diminuer le nombre de variations à interpréter et faciliter la découverte de gènes candidats.

La transformation tumorale étant un processus multifactoriel pouvant impliquer plusieurs événements moléculaires constitutionnels ou somatiques et des facteurs environnementaux, il est à mon sens compliqué d'estimer si une variation identifiée chez un seul patient est la cause principale à l'origine du cancer, si elle fait partie d'un faisceau d'évènements moléculaires ayant individuellement un effet faible mais dont la combinaison aboutit à la transformation tumorale ou si elle est un élément modificateur venant accélérer le processus de tumorigénèse du à un autre événement somatique ou non identifié par l'ES. Il me semble donc indispensable de prioriser les gènes candidats et de chercher à identifier des réplifications ou de réaliser des études fonctionnelles poussées pour évaluer l'impact de la variation.

Cette approche de séquençage pangénomiques dans les phénotypes extrêmes de cancer, portée par l'équipe depuis plusieurs années dans le cadre de la transposition des modèles issus des maladies rares en oncogénétique a récemment été reprise au niveau national. En effet, les deux plateformes pilotes du plan France Médecine Génomique vont se mettre en place prochainement, et 12 préindications ont été retenues. Parmi ces préindications figurent les patients atteints de cancers dans un contexte d'antécédents familiaux sévères évocateurs de prédispositions et pour lesquels les analyses en panel de gènes sont normales, ainsi que les cas extrêmes sporadiques de cancer, sujets concernés par le projet EXTRICAN. Seule une lecture diagnostique sera faite car ces plateformes n'ont pas directement de missions de recherche. Cela montre l'intérêt de l'approche tout en gardant l'importance du projet EXTRICAN qui a aussi pour but l'identification de nouveaux gènes candidats.

Bibliographie

- Abel, F., Sjöberg, R.-M., Ejeskär, K., Krona, C., and Martinsson, T. (2002). Analyses of apoptotic regulators CASP9 and DFFA at 1P36.2, reveal rare allele variants in human neuroblastoma tumours. *Br J Cancer* *86*, 596–604.
- Abel, F., Sjöberg, R.-M., Krona, C., Nilsson, S., and Martinsson, T. (2004). Mutations in the N-terminal domain of DFF45 in a primary germ cell tumor and in neuroblastoma tumors. *International Journal of Oncology* *25*, 1297–1302.
- Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. *Genome Biol* *17*.
- AlDubayan, S.H., Giannakis, M., Moore, N.D., Han, G.C., Reardon, B., Hamada, T., Mu, X.J., Nishihara, R., Qian, Z., Liu, L., et al. (2018). Inherited DNA-Repair Defects in Colorectal Cancer. *The American Journal of Human Genetics* *102*, 401–414.
- Alimonti, A., Carracedo, A., Clohessy, J.G., Trotman, L.C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W.J., Brogi, E., et al. (2010). Subtle variations in Pten dose determine cancer susceptibility. *Nature Genetics* *42*, 454–458.
- Anderson, W.F., Jatoi, I., Tse, J., and Rosenberg, P.S. (2010). Male Breast Cancer: A Population-Based Comparison With Female Breast Cancer. *Journal of Clinical Oncology* *28*, 232–239.
- Antczak, A., Kluźniak, W., Wokołorczyk, D., Kashyap, A., Jakubowska, A., Gronwald, J., Huzarski, T., Byrski, T., Dębniak, T., Masojć, B., et al. (2013). A common nonsense mutation of the BLM gene and prostate cancer risk and survival. *Gene* *532*, 173–176.
- Antoniou, A.C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkäs, K., Roberts, J., Lee, A., Subramanian, D., De Leeneer, K., Fostira, F., et al. (2014). Breast-Cancer Risk in Families with Mutations in PALB2. *New England Journal of Medicine* *371*, 497–506.
- Arena, Martini Maurizio, Luongo Myriam, Capelli Arnaldo, and Larocca Luigi Maria (2003). Mutations of the BIK gene in human peripheral B-cell lymphomas. *Genes, Chromosomes and Cancer* *38*, 91–96.
- Armitage, P., and Doll, R. (2004). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* *91*, 1983–1989.
- van Asperen, C.J., Brohet, R.M., Meijers-Heijboer, E.J., Hoogerbrugge, N., Verhoef, S., Vasen, H.F.A., Ausems, M.G.E.M., Menko, F.H., Gomez Garcia, E.B., Klijn, J.G.M., et al. (2005). Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *J. Med. Genet.* *42*, 711–719.
- Bachet, J.-B., Landi, B., Laurent-Puig, P., Italiano, A., Le Cesne, A., Lévy, P., Safar, V., Duffaud, F., Blay, J.-Y., and Emile, J.-F. (2013). Diagnosis, prognosis and treatment of patients with gastrointestinal stromal tumour (GIST) and germline mutation of KIT exon 13. *Eur. J. Cancer* *49*, 2531–2541.
- Bakker, J.L., Thirthagiri, E., Mil, S.E. van, Adank, M.A., Ikeda, H., Verheul, H.M.W., Meijers-Heijboer, H., Winter, J.P. de, Sharan, S.K., and Waisfisz, Q. (2014). A Novel Splice Site Mutation in the Noncoding Region of BRCA2: Implications for Fanconi Anemia and Familial Breast Cancer Diagnostics. *Human Mutation* *35*, 442–446.

- Basham, V.M., Lipscombe, J.M., Ward, J.M., Gayther, S.A., Ponder, B.A., Easton, D.F., and Pharoah, P.D. (2001). BRCA1 and BRCA2 mutations in a population-based study of male breast cancer. *Breast Cancer Research* 4.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59.
- Berger, A.H., Knudson, A.G., and Pandolfi, P.P. (2011). A continuum model for tumour suppression. *Nature* 476, 163–169.
- Bieche, I., Khodja, A., and Lidereau, R. (1998). Deletion mapping in breast tumor cell lines points to two distinct tumor-suppressor genes in the 1p32-pter region, one of deleted regions (1p36.2) being located within the consensus region of LOH in neuroblastoma. *Oncology Reports* 5, 267–339.
- Blue, G.M., Kirk, E.P., Giannoulatou, E., Dunwoodie, S.L., Ho, J.W.K., Hilton, D.C.K., White, S.M., Sholler, G.F., Harvey, R.P., and Winlaw, D.S. (2014). Targeted Next-Generation Sequencing Identifies Pathogenic Variants in Familial Congenital Heart Disease. *Journal of the American College of Cardiology* 64, 2498–2506.
- Bogdanova, N., Togo, A.V., Ratajska, M., Kluźniak, W., Takhirova, Z., Tarp, T., Prokofyeva, D., Bermisheva, M., Yanus, G.A., Gorodnova, T.V., et al. (2015). Prevalence of the BLM nonsense mutation, p.Q548X, in ovarian cancer patients from Central and Eastern Europe. *Familial Cancer* 14, 145–149.
- Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* 33, 228.
- Breast Cancer Linkage Consortium, T. (1999). Cancer Risks in BRCA2 Mutation Carriers. *JNCI Journal of the National Cancer Institute* 91, 1310–1316.
- Bredel, M., Bredel, C., Juric, D., Harsh, G.R., Vogel, H., Recht, L.D., and Sikic, B.I. (2005). High-Resolution Genome-Wide Mapping of Genetic Alterations in Human Glial Brain Tumors. *Cancer Res* 65, 4088–4096.
- Brinton, L.A., Key, T.J., Kolonel, L.N., Michels, K.B., Sesso, H.D., Ursin, G., Van Den Eeden, S.K., Wood, S.N., Falk, R.T., Parisi, D., et al. (2015). Prediagnostic Sex Steroid Hormones in Relation to Male Breast Cancer Risk. *Journal of Clinical Oncology* 33, 2041–2050.
- Brough, R., Frankum, J.R., Costa-Cabral, S., Lord, C.J., and Ashworth, A. (2011). Searching for synthetic lethality in cancer. *Current Opinion in Genetics & Development* 21, 34–41.
- Brown, E.J. (2003). Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes & Development* 17, 615–628.
- Brown, E.J., and Baltimore, D. (2000). ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* 14, 397–402.
- Brustmann, H. (2007). Poly(ADP-ribose) polymerase (PARP) and DNA-fragmentation factor (DFF45): Expression and correlation in normal, hyperplastic and neoplastic endometrial tissues. *Pathology - Research and Practice* 203, 65–72.

Byrski, T., Dent, R., Blecharz, P., Foszczynska-Kloda, M., Gronwald, J., Huzarski, T., Cybulski, C., Marczyk, E., Chrzan, R., Eisen, A., et al. (2012). Results of a phase II open-label, non-randomized trial of cisplatin chemotherapy in patients with BRCA1 -positive metastatic breast cancer. *Breast Cancer Research* 14, R110.

Byun, T.S. (2005). Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes & Development* 19, 1040–1052.

Cai, J., Feng, D., Hu, L., Chen, H., Yang, G., Cai, Q., Gao, C., and Wei, D. (2015). FAT4 functions as a tumour suppressor in gastric cancer by modulating Wnt/ β -catenin signalling. *British Journal of Cancer* 113, 1720.

Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.-L., et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467, 1109.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.

Casper, A.M., Nghiem, P., Arlt, M.F., and Glover, T.W. (2002). ATR regulates fragile site stability. *Cell* 111, 779–789.

Castells, A., Ino, Y., Louis, D.N., Ramesh, V., Gusella, J.F., and Rustgi, A.K. (1999). Mapping of a target region of allelic loss to a 0.5-cm interval on chromosome 22q13 in human colorectal cancer. *Gastroenterology* 117, 831–837.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* 2, 401–404.

Certo, M., Del Gaizo Moore, V., Nishino, M., Wei, G., Korsmeyer, S., Armstrong, S.A., and Letai, A. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* 9, 351–365.

Chakravarty, G., Hadsell, D., Buitrago, W., Settleman, J., and Rosen, J.M. (2003). p190-B RhoGAP regulates mammary ductal morphogenesis. *Mol. Endocrinol.* 17, 1054–1065.

Cheng, E.H.-Y.A., Wei, M.C., Weiler, S., Flavell, R.A., Mak, T.W., Lindsten, T., and Korsmeyer, S.J. (2001). BCL-2, BCL-XL Sequester BH3 Domain-Only Molecules Preventing BAX- and BAK-Mediated Mitochondrial Apoptosis. *Molecular Cell* 8, 705–711.

Chubb, D., Broderick, P., Dobbins, S.E., Frampton, M., Kinnersley, B., Penegar, S., Price, A., Ma, Y.P., Sherborne, A.L., Palles, C., et al. (2016). Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *Nature Communications* 7, 11883.

Ciriello, G., Cerami, E., Sander, C., and Schultz, N. (2012). Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* 22, 398–406.

Comings, D.E. (1973). A General Theory of Carcinogenesis. *Proc Natl Acad Sci U S A* 70, 3324–3328.

Comino-Méndez, I., Gracia-Aznárez, F.J., Schiavi, F., Landa, I., Leandro-García, L.J., Letón, R., Honrado, E., Ramos-Medina, R., Caronia, D., Pita, G., et al. (2011). Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nature Genetics* 43, 663–667.

- Cortez, D., Guntuku, S., Qin, J., and Elledge, S.J. (2001). ATR and ATRIP: partners in checkpoint signaling. *Science* 294, 1713–1716.
- Couch, F.J., Shimelis, H., Hu, C., Hart, S.N., Polley, E.C., Na, J., Hallberg, E., Moore, R., Thomas, A., Lilyquist, J., et al. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol* 3, 1190–1196.
- Curtin, N.J. (2012). DNA repair dysregulation from cancer driver to therapeutic target. *Nature Reviews Cancer* 12, 801.
- Dai, Z., Liu, S., Marcucci, G., and Sadee, W. (2006). 5-Aza-2'-deoxycytidine and depsipeptide synergistically induce expression of BIK (BCL2-interacting killer). *Biochemical and Biophysical Research Communications* 351, 455–461.
- Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46, D794–D801.
- Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., and Bérout, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37, e67–e67.
- Dragojlovic, N., Elliott, A.M., Adam, S., Karnebeek, C. van, Lehman, A., Mwenifumbo, J.C., Nelson, T.N., Souich, C. du, Friedman, J.M., and Lynd, L.D. (2018). The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. *Genetics in Medicine* 20, 1013.
- Duns, G., Berg, E. van den, Duivenbode, I. van, Osinga, J., Hollema, H., Hofstra, R.M.W., and Kok, K. (2010). Histone Methyltransferase Gene SETD2 Is a Novel Tumor Suppressor Gene in Clear Cell Renal Cell Carcinoma. *Cancer Res* 70, 4287–4291.
- Durkin, M.E., Yuan, B.-Z., Zhou, X., Zimonjic, D.B., Lowy, D.R., Thorgeirsson, S.S., and Popescu, N.C. (2007). DLC-1: a Rho GTPase-activating protein and tumour suppressor. *Journal of Cellular and Molecular Medicine* 11, 1185.
- Durocher, F., Labrie, Y., Soucy, P., Sinilnikova, O., Labuda, D., Bessette, P., Chiquette, J., Laframboise, R., Lépine, J., Lespérance, B., et al. (2006). Mutation analysis and characterization of ATR sequence variants in breast cancer cases from high-risk French Canadian breast/ovarian cancer families. *BMC Cancer* 6, 230.
- Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO Journal* 27, 406.
- Edwards, S.M., Kote-Jarai, Z., Meitz, J., Hamoudi, R., Hope, Q., Osin, P., Jackson, R., Southgate, C., Singh, R., Falconer, A., et al. (2003). Two percent of men with early-onset prostate cancer harbor germline mutations in the BRCA2 gene. *Am. J. Hum. Genet.* 72, 1–12.
- ERISCAM (2010). Estimation des risques tumoraux dans le syndrome de Lynch – Résultats de l'étude française ERISCAM – Mise à Jour 2010 | hnpcc-lynch.
- Errami, Y., Brim, H., Oumouna-Benachour, K., Oumouna, M., Naura, A.S., Kim, H., Ju, J., Davis, C.J., Kim, J.G., Ashktorab, H., et al. (2013). ICAD Deficiency in Human Colon Cancer and Predisposition to Colon Tumorigenesis: Linkage to Apoptosis Resistance and Genomic Instability. *PLOS ONE* 8, e57871.

Espenschied, C.R., LaDuca, H., Li, S., McFarland, R., Gau, C.-L., and Hampel, H. (2017). Multigene Panel Testing Provides a New Perspective on Lynch Syndrome. *JCO* 35, 2568–2575.

Evan, G.I., Wyllie, A.H., Gilbert, C.S., Littlewood, T.D., Land, H., Brooks, M., Waters, C.M., Penn, L.Z., and Hancock, D.C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell* 69, 119–128.

Evans, D.G.R., Susnerwala, I., Dawson, J., Woodward, E., Maher, E.R., and Lalloo, F. (2010). Risk of breast cancer in male BRCA2 carriers. *Journal of Medical Genetics* 47, 710–711.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434, 917.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res* 34, e22–e22.

Fitzgerald, T.W., Gerety, S.S., Jones, W.D., Kogelenberg, M. van, King, D.A., McRae, J., Morley, K.I., Parthiban, V., Al-Turki, S., Ambridge, K., et al. (2015). Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 519, 223.

Flatt, P.M., Polyak, K., Tang, L.J., Scatena, C.D., Westfall, M.D., Rubinstein, L.A., Yu, J., Kinzler, K.W., Vogelstein, B., Hill, D.E., et al. (2000). p53-dependent expression of PIG3 during proliferation, genotoxic stress, and reversible growth arrest. *Cancer Lett.* 156, 63–72.

Fletcher, O., and Houlston, R.S. (2010). Architecture of inherited susceptibility to common cancer. *Nature Reviews Cancer* 10, 353.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers.

Fontebasso, A.M., Schwartzenruber, J., Khuong-Quang, D.-A., Liu, X.-Y., Sturm, D., Korshunov, A., Jones, D.T.W., Witt, H., Kool, M., Albrecht, S., et al. (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol* 125, 659–669.

Foulkes, W.D. (2008). Inherited Susceptibility to Common Cancers. *New England Journal of Medicine* 359, 2143–2153.

Foulkes, W.D., Ghadirian, P., Akbari, M.R., Hamel, N., Giroux, S., Sabbaghian, N., Darnel, A., Royer, R., Poll, A., Fafard, E., et al. (2007). Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Research* 9, R83.

Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth. *The American Journal of Human Genetics* 91, 597–607.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nature Reviews Cancer* 4, 177–183.

- Garber, J.E., and Offit, K. (2005). Hereditary Cancer Predisposition Syndromes. *JCO* 23, 276–292.
- Garcia-Muse, T., and Boulton, S.J. (2005). Distinct modes of ATR activation after replication stress and DNA double-strand breaks in *Caenorhabditis elegans*. *The EMBO Journal* 24, 4345–4355.
- Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing.
- GGC-Unicancer (2017). UNICANCER - Quels gènes analyser face à un risque de cancer héréditaire du sein et de l'ovaire ?
- Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., Vorst, M. van de, Bon, B.W.M. van, Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347.
- Glenn, T.C. (2011). Field guide to next-generation DNA sequencers. *Mol Ecol Resour* 11, 759–769.
- Goss, K.H., Risinger, M.A., Kordich, J.J., Sanz, M.M., Straughen, J.E., Slovek, L.E., Capobianco, A.J., German, J., Boivin, G.P., and Groden, J. (2002). Enhanced Tumor Formation in Mice Heterozygous for Blm Mutation. *Science* 297, 2051–2053.
- Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N.D., Kanchi, K.L., Maher, C.A., Fulton, R., Fulton, L., Wallis, J., et al. (2012). Genomic Landscape of Non-Small Cell Lung Cancer in Smokers and Never-Smokers. *Cell* 150, 1121–1134.
- Greenleaf, W.J., and Sidow, A. (2014). The future of sequencing: convergence of intelligent design and market Darwinism. *Genome Biology* 15, 303.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.
- Gumaste, P.V., Penn, L.A., Cymerman, R.M., Kirchhoff, T., Polsky, D., and McLellan, B. (2015). Skin cancer risk in BRCA1/2 mutation carriers. *Br. J. Dermatol.* 172, 1498–1506.
- Half, E., Bercovich, D., and Rozen, P. (2009). Familial adenomatous polyposis. *Orphanet Journal of Rare Diseases* 4, 22.
- Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. *Cell* 100, 57–70.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.
- Hayes, J.L., Tzika, A., Thygesen, H., Berri, S., Wood, H.M., Hewitt, S., Pendlebury, M., Coates, A., Willoughby, L., Watson, C.M., et al. (2013). Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics* 102, 174–181.
- Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1–8.

Heckman-Stoddard, B.M., Vargo-Gogola, T., Herrick, M.P., Visbal, A.P., Lewis, M.T., Settleman, J., and Rosen, J.M. (2011). P190A RhoGAP is required for mammary gland development. *Dev. Biol.* 360, 1–10.

Helbig, K.L., Hagman, K.D.F., Shinde, D.N., Mroske, C., Powis, Z., Li, S., Tang, S., and Helbig, I. (2016). Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. *Genetics in Medicine* 18, 898.

Hinck, L., and Silberstein, G.B. (2005). Key stages in mammary gland development: the mammary end bud as a motile organ. *Breast Cancer Res.* 7, 245–251.

Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS* 106, 9362–9367.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., et al. (2013). TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science* 339, 959–961.

Hou, L., Chen, M., Zhao, X., Li, J., Deng, S., Hu, J., Yang, H., and Jiang, J. (2016). FAT4 functions as a tumor suppressor in triple-negative breast cancer. *Tumor Biol.* 37, 16337–16343.

Hur, J., Chesnes, J., Coser, K.R., Lee, R.S., Geck, P., Isselbacher, K.J., and Shioda, T. (2004). The BIK BH3-only protein is induced in estrogen-starved and antiestrogen-exposed breast cancer cells and provokes apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2351–2356.

Hur, J., Bell, D.W., Dean, K.L., Coser, K.R., Hilario, P.C., Okimoto, R.A., Tobey, E.M., Smith, S.L., Isselbacher, K.J., and Shioda, T. (2006). Regulation of Expression of BIK Proapoptotic Protein in Human Breast Cancer Cells: p53-Dependent Induction of BIK mRNA by Fulvestrant and Proteasomal Degradation of BIK Protein. *Cancer Res* 66, 10153–10161.

Iglesias, A., Anyane-Yeboah, K., Wynn, J., Wilson, A., Cho, M.T., Guzman, E., Sisson, R., Egan, C., and Chung, W.K. (2014). The usefulness of whole-exome sequencing in routine clinical practice. *Genetics in Medicine* 16, 922.

INCa (2015a). *Epidémiologie des cancers - Les chiffres du cancer en France | Institut National Du Cancer.*

INCa (2015b). *Les prédispositions génétiques - Oncogénétique | Institut National Du Cancer.*

INCa (2016). *Oncogénétique en 2016 - Consultations, laboratoires et suivi - Ref : ADONCOG17 | Institut National Du Cancer.*

INCa (2017). *Cancer de l’ovaire et inhibiteur de PARP - Oncogénétique et plateformes de génétique moléculaire.*

Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216.

Iyer, D.R., and Rhind, N. (2017). The Intra-S Checkpoint Responses to DNA Damage. *Genes* 8, 74.

- Jiang, Y., Shan, S., Chi, L., Zhang, G., Gao, X., Li, H., Zhu, X., and Yang, J. (2016). Methyl methanesulfonate induces necroptosis in human lung adenoma A549 cells through the PIG-3-reactive oxygen species pathway. *Tumor Biol.* *37*, 3785–3795.
- Jones, S., Hruban, R.H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D.W., Lin, J.C.-H., Palmisano, E., Brune, K., Jaffee, E.M., et al. (2009). Exomic Sequencing Identifies PALB2 as a Pancreatic Cancer Susceptibility Gene. *Science* *324*, 217–217.
- Jorquera, R., and Tanguay, R.M. (2001). Fumarylacetoacetate, the metabolite accumulating in hereditary tyrosinemia, activates the ERK pathway and induces mitotic abnormalities and genomic instability. *Hum. Mol. Genet.* *10*, 1741–1752.
- Józwiak, S., Stein, K., and Kotulska, K. (2012). Everolimus (RAD001): first systemic treatment for subependymal giant cell astrocytoma associated with tuberous sclerosis complex. *Future Oncol* *8*, 1515–1523.
- Kanu, N., Grönroos, E., Martinez, P., Burrell, R.A., Goh, X.Y., Bartkova, J., Maya-Mendoza, A., Mistrík, M., Rowan, A.J., Patel, H., et al. (2015). SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. *Oncogene* *34*, 5699.
- Katoh, M. (2012). Function and cancer genomics of FAT family genes (review). - PubMed - NCBI.
- Kernohan, K.D., Dymont, D.A., Pupavac, M., Cramer, Z., McBride, A., Bernard, G., Straub, I., Tetreault, M., Hartley, T., Huang, L., et al. (2017). Matchmaking facilitates the diagnosis of an autosomal-recessive mitochondrial disease caused by biallelic mutation of the tRNA isopentenyltransferase (TRIT1) gene. *Human Mutation* *38*, 511–516.
- Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M.A., and Gerstein, M. (2016). Role of non-coding sequence variants in cancer. *Nature Reviews Genetics* *17*, 93–108.
- Kinzler, K.W., and Vogelstein, B. (1997). Gatekeepers and caretakers. *Nature* *386*, 761–763.
- de Klein, A., Muijtjens, M., van Os, R., Verhoeven, Y., Smit, B., Carr, A.M., Lehmann, A.R., and Hoeijmakers, J.H.J. (2000). Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. *Current Biology* *10*, 479–482.
- Knudson, A.G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc Natl Acad Sci U S A* *68*, 820–823.
- Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell* *155*, 27–38.
- Kodera, H., Ohba, C., Kato, M., Maeda, T., Araki, K., Tajima, D., Matsuo, M., Hino-Fukuyo, N., Kohashi, K., Ishiyama, A., et al. (2016). De novo GABRA1 mutations in Ohtahara and West syndromes. *Epilepsia* *57*, 566–573.
- Konishi, S., Ishiguro, H., Shibata, Y., Kudo, J., Terashita, Y., Sugiura, H., Koyama, H., Kimura, M., Sato, A., Shinoda, N., et al. (2002). Decreased expression of DFF45/ICAD is correlated with a poor prognosis in patients with esophageal carcinoma. *Cancer* *95*, 2473–2478.
- Korde, L.A., Zujewski, J.A., Kamin, L., Giordano, S., Domchek, S., Anderson, W.F., Bartlett, J.M.S., Gelmon, K., Nahleh, Z., Bergh, J., et al. (2010). Multidisciplinary Meeting on Male Breast Cancer: Summary and Research Recommendations. *JCO* *28*, 2114–2122.

- Kovacs, G., Szücs, S., Riese, W.D., and Baumgärtel, H. (1987). Specific chromosome aberration in human renal cell carcinoma. *International Journal of Cancer* *40*, 171–178.
- Krumm, N., Sudmant, P.H., Ko, A., O’Roak, B.J., Malig, M., Coe, B.P., Project, N.E.S., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Res.* *22*, 1525–1532.
- Kuchenbaecker, K.B., Hopper, J.L., Barnes, D.R., Phillips, K.-A., Mooij, T.M., Roos-Blom, M.-J., Jervis, S., Leeuwen, F.E. van, Milne, R.L., Andrieu, N., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* *317*, 2402–2416.
- Kunkel, T.A. (2009). Evolving Views of DNA Replication (In)Fidelity. *Cold Spring Harb Symp Quant Biol* *74*, 91–101.
- Lee, A.J., Cunningham, A.P., Kuchenbaecker, K.B., Mavaddat, N., Easton, D.F., and Antoniou, A.C. (2014). BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer* *110*, 535–545.
- Lee, J.-H., Kang, Y., Khare, V., Jin, Z.-Y., Kang, M.-Y., Yoon, Y., Hyun, J.-W., Chung, M.-H., Cho, S.-I., Jun, J.Y., et al. (2010). The p53-inducible gene 3 (PIG3) contributes to early cellular response to DNA damage. *Oncogene* *29*, 1431–1450.
- Lejeune, F., and Maquat, L.E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current Opinion in Cell Biology* *17*, 309–315.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O’Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285–291.
- Lempiäinen, H., and Halazonetis, T.D. (2009). Emerging common themes in regulation of PIKKs and PI3Ks. *EMBO J* *28*, 3067–3073.
- Letai, A. (2017). Apoptosis and Cancer. *Annual Review of Cancer Biology* *1*, 275–294.
- Letai, A., Bassik, M.C., Walensky, L.D., Sorcinelli, M.D., Weiler, S., and Korsmeyer, S.J. (2002). Distinct BH3 domains either sensitize or activate mitochondrial apoptosis, serving as prototype cancer therapeutics. *Cancer Cell* *2*, 183–192.
- Lewis, K.A., Mullany, S., Thomas, B., Chien, J., Loewen, R., Shridhar, V., and Cliby, W.A. (2005). Heterozygous ATR mutations in mismatch repair-deficient cancer cells have functional significance. *Cancer Res.* *65*, 7091–7095.
- Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.-M. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutS α . *Cell* *153*, 590–600.
- Li, J., Duns, G., Westers, H., Sijmons, R., Berg, A. van den, Kok, K., Li, J., Duns, G., Westers, H., Sijmons, R., et al. (2016a). SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget* *7*, 50719–50734.
- Li, J., Kluiver, J., Osinga, J., Westers, H., van Werkhoven, M.B., Seelen, M.A., Sijmons, R.H., van den Berg, A., and Kok, K. (2016b). Functional Studies on Primary Tubular Epithelial Cells Indicate a Tumor Suppressor Role of SETD2 in Clear Cell Renal Cell Carcinoma. *Neoplasia* *18*, 339–346.

- Lièvre, A., and Laurent-Puig, P. (2010). RAS/MAPK signaling pathway.
- Liu, A., Takakuwa, T., Fujita, S., Ham, M.F., Luo, W.-J., Daibata, M., and Aozasa, K. (2005). Alterations of DNA damage-response genes ATM and ATR in pyothorax-associated lymphoma. *Lab. Invest.* *85*, 436–446.
- Liu, A., Takakuwa, T., Fujita, S., Luo, W.-J., Tresnasari, K., Van den Berg, A., Poppema, S., and Aozasa, K. (2008). ATR alterations in Hodgkin's lymphoma. *Oncology Reports*.
- Ljungman, M., and Lane, D.P. (2004). Transcription — guarding the genome by sensing DNA damage. *Nature Reviews Cancer* *4*, 727–737.
- Logan, T.F. (2013). Foretinib (XL880): c-MET Inhibitor with Activity in Papillary Renal Cell Cancer. *Curr Oncol Rep* *15*, 83–90.
- Long, P.A., Larsen, B.T., Evans, J.M., and Olson, T.M. (2015). Exome Sequencing Identifies Pathogenic and Modifier Mutations in a Child With Sporadic Dilated Cardiomyopathy. *Journal of the American Heart Association*.
- Lukasik, D., Wilczek, E., Wasiutynski, A., and Gornicka, B. (2011). Deleted in liver cancer protein family in human malignancies (Review). *Oncology Letters* *2*, 763.
- Ly, D., Forman, D., Ferlay, J., Brinton, L.A., and Cook, M.B. (2013). An international comparison of male and female breast cancer incidence rates. *International Journal of Cancer* *132*, 1918–1926.
- Majewski, J., Schwartzenuber, J., Lalonde, E., Montpetit, A., and Jabado, N. (2011). What can exome sequencing do for you? *Journal of Medical Genetics* *48*, 580–589.
- Mathai, J.P., Germain, M., Marcellus, R.C., and Shore, G.C. (2002). Induction and endoplasmic reticulum location of BIK/NBK in response to apoptotic signaling by E1A and p53. *Oncogene* *21*, 2534.
- Mathai, J.P., Germain, M., and Shore, G.C. (2005). BH3-only BIK Regulates BAX,BAK-dependent Release of Ca²⁺ from Endoplasmic Reticulum Stores and Mitochondrial Apoptosis during Stress-induced Cell Death. *J. Biol. Chem.* *280*, 23829–23836.
- Mazouzi, A., Velimezi, G., and Loizou, J.I. (2014). DNA replication stress: Causes, resolution and disease. *Experimental Cell Research* *329*, 85–93.
- McCormack, J., Welsh, N.J., and Braga, V.M.M. (2013). Cycling around cell-cell adhesion with Rho GTPase regulators. *J. Cell. Sci.* *126*, 379–391.
- Medema, R.H., and Macůrek, L. (2012). Checkpoint control and cancer. *Oncogene* *31*, 2601–2613.
- Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., et al. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat. Genet.* *31*, 55–59.
- Menoyo, A., Alazzouzi, H., Espín, E., Armengol, M., Yamamoto, H., and Schwartz, S. (2001). Somatic Mutations in the DNA Damage-Response Genes ATR and CHK1 in Sporadic Stomach Tumors with Microsatellite Instability. *Cancer Res* *61*, 7727–7730.

- Metzker, M.L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics* *11*, 31–46.
- Mokrani-Benhelli, H., Gaillard, L., Biasutto, P., Le Guen, T., Touzot, F., Vasquez, N., Komatsu, J., Conseiller, E., Picard, C., Gluckman, E., et al. (2013). Primary Microcephaly, Impaired DNA Replication, and Genomic Instability Caused by Compound Heterozygous ATR Mutations. *Human Mutation* *34*, 374–384.
- Moll, U.M., and Petrenko, O. (2003). The MDM2-p53 Interaction. *Mol Cancer Res* *1*, 1001–1008.
- Morris, A.P., Tanya M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* *44*, 981.
- Murphy, K.M., Brune, K.A., Griffin, C., Sollenberger, J.E., Petersen, G.M., Bansal, R., Hruban, R.H., and Kern, S.E. (2002). Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. *Cancer Res.* *62*, 3789–3793.
- Nagata, S. (2000). Apoptotic DNA Fragmentation. *Experimental Cell Research* *256*, 12–18.
- Nakagawara, A., Ohira, M., Kageyama, H., Mihara, M., Furuta, S., Machida, T., Takayasu, H., Islam, A., Nakamura, Y., Takahashi, M., et al. (2000). Identification of the homozygously deleted region at chromosome 1p36.2 in human neuroblastoma. *Medical and Pediatric Oncology* *35*, 516–521.
- Nambot, S., Thevenon, J., Kuentz, P., Duffourd, Y., Tisserant, E., Bruel, A.-L., Mosca-Boidron, A.-L., Masurel-Paulet, A., Lehalle, D., Jean-Marçais, N., et al. (2018). Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genetics in Medicine* *20*, 645–654.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Eposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* *472*, 90–94.
- Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics* *42*, 30.
- Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012). The Life History of 21 Breast Cancers. *Cell* *149*, 994–1007.
- Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., and Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* *253*, 665–669.
- O'Connor, C. (2008). Human Chromosome Translocations and Cancer | Learn Science at Scitable. Nature Education.
- O'Driscoll, M., and Jeggo, P.A. (2006). The role of double-strand break repair — insights from human genetics. *Nature Reviews Genetics* *7*, 45–54.

O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A., and Goodship, J.A. (2003). A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat. Genet.* *33*, 497–501.

Ogi, T., Walker, S., Stiff, T., Hobson, E., Limsirichaikul, S., Carpenter, G., Prescott, K., Suri, M., Byrd, P.J., Matsuse, M., et al. (2012). Identification of the First ATRIP-Deficient Patient and Novel Mutations in ATR Define a Clinical Spectrum for ATR-ATRIP Seckel Syndrome. *PLOS Genetics* *8*, e1002945.

Oh, M., McBride, A., Yun, S., Bhattacharjee, S., Slack, M., Martin, J.R., Jeter, J., and Abraham, I. (2018). BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. *J. Natl. Cancer Inst.* *110*, 1178–1189.

Ottini, L., Rizzolo, P., Zanna, I., Falchetti, M., Masala, G., Ceccarelli, K., Vezzosi, V., Gulino, A., Giannini, G., Bianchi, S., et al. (2009). BRCA1/BRCA2 mutation status and clinical-pathologic features of 108 male breast cancer cases from Tuscany: a population-based study in central Italy. *Breast Cancer Research and Treatment* *116*, 577–586.

Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics* *45*, 136–144.

Park, H., Cho, S.-Y., Kim, H., Na, D., Han, J.Y., Chae, J., Park, C., Park, O.-K., Min, S., Kang, J., et al. (2015). Genomic alterations in BCL2L1 and DLC1 contribute to drug sensitivity in gastric cancer. *PNAS* *112*, 12492–12497.

Park, I.Y., Powell, R.T., Tripathi, D.N., Dere, R., Ho, T.H., Blasius, T.L., Chiang, Y.-C., Davis, I.J., Fahey, C.C., Hacker, K.E., et al. (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell* *166*, 950–962.

Parsons, M.T., Whiley, P.J., Beesley, J., Drost, M., Wind, N. de, Thompson, B.A., Marquart, L., Hopper, J.L., Jenkins, M.A., Brown, M.A., et al. (2015). Consequences of germline variation disrupting the constitutional translational initiation codon start sites of MLH1 and BRCA2: Use of potential alternative start sites and implications for predicting variant pathogenicity. *Molecular Carcinogenesis* *54*, 513–522.

Pearlman, R., Frankel, W.L., Swanson, B., Zhao, W., Yilmaz, A., Miller, K., Bacher, J., Bigley, C., Nelsen, L., Goodfellow, P.J., et al. (2017). Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol* *3*, 464–471.

Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W., and Vogelstein, B. (1997). A model for p53-induced apoptosis. *Nature* *389*, 300–305.

Popescu, N.C., and Goodison, S. (2014). Deleted in Liver Cancer-1 (DLC1): An Emerging Metastasis Suppressor Gene. *Mol Diagn Ther* *18*, 293–302.

Porté, S., Valencia, E., Yakovtseva, E.A., Borràs, E., Shafiqat, N., Debreczeny, J.É., Pike, A.C.W., Oppermann, U., Farrés, J., Fita, I., et al. (2009). Three-dimensional Structure and Enzymatic Function of Proapoptotic Human p53-inducible Quinone Oxidoreductase PIG3. *J. Biol. Chem.* *284*, 17194–17205.

Powis, Z., Espenschied, C.R., LaDuca, H., Hagman, K.D., Paudyal, T., Li, S., Inaba, H., Mauer, A., Nathanson, K.L., Knost, J., et al. (2018). Clinical germline diagnostic exome sequencing for hereditary cancer: Findings within novel candidate genes are prevalent. *Cancer Genetics* 224, 12–20.

Prokofyeva, D., Bogdanova, N., Dubrowinskaja, N., Bermisheva, M., Takhirova, Z., Antonenkova, N., Turmanov, N., Datsyuk, I., Gantsev, S., Christiansen, H., et al. (2013). Nonsense mutation p.Q548X in BLM, the gene mutated in Bloom's syndrome, is associated with breast cancer in Slavic populations. *Breast Cancer Res Treat* 137, 533–539.

Qi, C., Zhu, Y.T., Hu, L., and Zhu, Y.-J. (2009). Identification of Fat4 as a candidate tumor suppressor gene in breast cancers. *Int J Cancer* 124, 793–798.

Quan, J., Li, Y., Jin, M., Chen, D., Yin, X., and Jin, M. (2017). Suppression of p53-inducible gene 3 is significant for glioblastoma progression and predicts poor patient prognosis. *Tumour Biol.* 39, 1010428317694572.

Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308.

Rahman, N., and Scott, R.H. (2007). Cancer genes associated with phenotypes in monoallelic and biallelic mutation carriers: new lessons from old players. *Hum Mol Genet* 16, R60–R66.

Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T., et al. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature Genetics* 39, 165–167.

Rauch, Wang, and Pfeifer (2012). DNA methylation biomarkers for lung cancer | SpringerLink.

Real, P.J., Sanz, C., Gutierrez, O., Pipaon, C., Zubiaga, A.M., and Fernandez-Luna, J.L. (2006). Transcriptional activation of the proapoptotic bik gene by E2F proteins in cancer cells. *FEBS Letters* 580, 5905–5909.

Redza-Dutordoir, M., and Averill-Bates, D.A. (2016). Activation of apoptosis signalling pathways by reactive oxygen species. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863, 2977–2992.

Reis, P.P., Rogatto, S.R., Kowalski, L.P., Nishimoto, I.N., Montovani, J.C., Corpus, G., Squire, J.A., and Kamel-Reid, S. (2002). Quantitative real-time PCR identifies a critical region of deletion on 22q13 related to prognosis in oral cancer. *Oncogene* 21, 6480.

Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., North, B., Jayatilake, H., Barfoot, R., Spanova, K., et al. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.* 38, 873–875.

Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* 18, 696–704.

Rizzolo, P., Silvestri, V., Tommasi, S., Pinto, R., Danza, K., Falchetti, M., Gulino, M., Frati, P., and Ottini, L. (2013). Male breast cancer: genetics, epigenetics, and ethical aspects. *Annals of Oncology* 24, viii75–viii82.

Roignot, J., Peng, X., and Mostov, K. (2013). Polarity in mammalian epithelial morphogenesis. *Cold Spring Harb Perspect Biol* 5.

- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996). Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry* 242, 84–89.
- Ronaghi, M., Uhlén, M., and Nyrén, P. (1998). A Sequencing Method Based on Real-Time Pyrophosphate. *Science* 281, 363–365.
- Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352.
- Rowley, J.D. (2001). Chromosome translocations: dangerous liaisons revisited. *Nature Reviews Cancer* 1, 245.
- Sahai, E., and Marshall, C.J. (2002). RHO–GTPases and cancer. *Nature Reviews Cancer* 2, 133.
- Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics* 46, 944–950.
- Sampson, J.R., Dolwani, S., Jones, S., Eccles, D., Ellis, A., Evans, D.G., Frayling, I., Jordan, S., Maher, E.R., Mak, T., et al. (2003). Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *The Lancet* 362, 39–41.
- Schwab, M., Praml, C., and Amler, L.C. (1996). Genomic instability in Ip and human malignancies. *Genes, Chromosomes and Cancer* 16, 211–229.
- Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell* 88, 593–602.
- Shah, S.P., Morin, R.D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., et al. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 461, 809–813.
- Shaheen, R., Patel, N., Shamseldin, H., Alzahrani, F., Al-Yamany, R., ALMoisheer, A., Ewida, N., Anazi, S., Alnemer, M., Elsheikh, M., et al. (2016). Accelerating matchmaking of novel dysmorphology syndromes through clinical and genomic characterization of a large cohort. *Genetics in Medicine* 18, 686.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* 309, 1728–1732.
- Shimada, M., Sagae, R., Kobayashi, J., Habu, T., and Komatsu, K. (2009). Inactivation of the Nijmegen Breakage Syndrome Gene Leads to Excess Centrosome Duplication via the ATR/BRCA1 Pathway. *Cancer Res* 69, 1768–1775.
- Silberstein, G.B. (2001). Postnatal mammary gland morphogenesis. *Microsc. Res. Tech.* 52, 155–162.
- Smith, M.J., O’Sullivan, J., Bhaskar, S.S., Hadfield, K.D., Poke, G., Caird, J., Sharif, S., Eccles, D., Fitzpatrick, D., Rawluk, D., et al. (2013). Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nature Genetics* 45, 295.

- Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum Mutat* 36, 928–930.
- Sokolenko, A.P., Iyevleva, A.G., Preobrazhenskaya, E.V., Mitiushkina, N.V., Abysheva, S.N., Suspitsin, E.N., Kuligina, E.Sh., Gorodnova, T.V., Pfeifer, W., Togo, A.V., et al. (2012). High prevalence and breast cancer predisposing role of the BLM c.1642 C>T (Q548X) mutation in Russia. *Int. J. Cancer* 130, 2867–2873.
- Sopik, V., Phelan, C., Cybulski, C., and Narod, S.A. (2015). BRCA1 and BRCA2 mutations and the risk for colorectal cancer. *Clin. Genet.* 87, 411–418.
- Sottoriva, A., Spiteri, I., Shibata, D., Curtis, C., and Tavaré, S. (2013). Single-Molecule Genomic Data Delineate Patient-Specific Tumor Profiles and Cancer Stem Cell Organization. *Cancer Res* 73, 41–49.
- Spurrell, C.H. (2013). Identifying New Genes for Inherited Breast Cancer by Exome Sequencing. University of Washington.
- Stadler, Z.K., Thom, P., Robson, M.E., Weitzel, J.N., Kauff, N.D., Hurley, K.E., Devlin, V., Gold, B., Klein, R.J., and Offit, K. (2010). Genome-Wide Association Studies of Cancer. *JCO* 28, 4255–4267.
- Stafford, J.L., Dyson, G., Levin, N.K., Chaudhry, S., Rosati, R., Kalpage, H., Wernette, C., Petrucelli, N., Simon, M.S., and Tainsky, M.A. (2017). Reanalysis of BRCA1/2 negative high risk ovarian cancer patients reveals novel germline risk loci and insights into missing heritability. *PLOS ONE* 12, e0178450.
- Stein, L.D. (2010). The case for cloud computing in genome informatics. *Genome Biology* 11, 207.
- Stephens, P.J., Tarpey, P.S., Davies, H., Loo, P.V., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486, 400–404.
- Stessman, H.A.F., Willemsen, M.H., Fenckova, M., Penn, O., Hoischen, A., Xiong, B., Wang, T., Hoekzema, K., Vives, L., Vogel, I., et al. (2016). Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. *The American Journal of Human Genetics* 98, 541–552.
- Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., et al. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and Cellular Biology* 22, 1298–1306.
- Sturm, I., Stephan, C., Gillissen, B., Siebert, R., Janz, M., Radetzki, S., Jung, K., Loening, S., Dörken, B., and Daniel, P.T. (2005). Loss of the tissue-specific proapoptotic BH3-only protein Nbk/Bik is a unifying feature of renal cell carcinoma. *Cell Death Differ* 13, 619–627.
- Suspitsin, E.N., Yanus, G.A., Sokolenko, A.P., Yatsuk, O.S., Zaitseva, O.A., Bessonov, A.A., Ivantsov, A.O., Heinstein, V.A., Klimashevskiy, V.F., Togo, A.V., et al. (2014). Development of breast tumors in CHEK2, NBN/NBS1 and BLM mutation carriers does not commonly involve somatic inactivation of the wild-type allele. *Med. Oncol.* 31, 828.
- Susswein, L.R., Marshall, M.L., Nusbaum, R., Vogel Postula, K.J., Weissman, S.M., Yackowski, L., Vaccari, E.M., Bissonnette, J., Booker, J.K., Cremona, M.L., et al. (2016). Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genetics in Medicine* 18, 823–832.

- Tai, Y.C., Domchek, S., Parmigiani, G., and Chen, S. (2007). Breast Cancer Risk Among Male BRCA1 and BRCA2 Mutation Carriers. *JNCI Journal of the National Cancer Institute* 99, 1811–1814.
- Tanaka, A., Weinel, S., Nagy, N., O’Driscoll, M., Lai-Cheong, J.E., Kulp-Shorten, C.L., Knable, A., Carpenter, G., Fisher, S.A., Hiragun, M., et al. (2012). Germline mutation in ATR in autosomal-dominant oropharyngeal cancer syndrome. *Am. J. Hum. Genet.* 90, 511–517.
- Tang, J.Y., Mackay-Wiggan, J.M., Aszterbaum, M., Yauch, R.L., Lindgren, J., Chang, K., Coppola, C., Chanana, A.M., Marji, J., Bickers, D.R., et al. (2012). Inhibiting the hedgehog pathway in patients with the basal-cell nevus syndrome. *N. Engl. J. Med.* 366, 2180–2188.
- Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707.
- Thauvin-Robinet, C., Auclair, M., Duplomb, L., Caron-Debarle, M., Avila, M., St-Onge, J., Le Merrer, M., Le Luyer, B., Héron, D., Mathieu-Dramard, M., et al. (2013). PIK3R1 mutations cause syndromic insulin resistance with lipoatrophy. *Am. J. Hum. Genet.* 93, 141–149.
- Thauvin-Robinet, C., Lee, J.S., Lopez, E., Herranz-Pérez, V., Shida, T., Franco, B., Jegou, L., Ye, F., Pasquier, L., Loget, P., et al. (2014). The oral-facial-digital syndrome gene C2CD3 encodes a positive regulator of centriole elongation. *Nat. Genet.* 46, 905–911.
- Thomas, C., and Strutt, D. (2012). The roles of the cadherins Fat and Dachsous in planar polarity specification in *Drosophila*. *Developmental Dynamics* 241, 27–39.
- Thompson, D., and Easton, D.F. (2002). Cancer Incidence in BRCA1 Mutation Carriers. *Journal of the National Cancer Institute* 94, 8.
- Thompson, E.R., Doyle, M.A., Ryland, G.L., Rowley, S.M., Choong, D.Y.H., Tothill, R.W., Thorne, H., kConFab, Barnes, D.R., Li, J., et al. (2012). Exome Sequencing Identifies Rare Deleterious Mutations in DNA Repair Genes FANCC and BLM as Potential Breast Cancer Susceptibility Alleles. *PLoS Genetics* 8, e1002894.
- Thompson, E.R., Rowley, S.M., Li, N., McInerney, S., Devereux, L., Wong-Brown, M.W., Trainer, A.H., Mitchell, G., Scott, R.J., James, P.A., et al. (2016). Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. *Journal of Clinical Oncology* 34, 1455–1459.
- Tomasetti, C., Marchionni, L., Nowak, M.A., Parmigiani, G., and Vogelstein, B. (2015). Only three driver gene mutations are required for the development of lung and colorectal cancers. *PNAS* 112, 118–123.
- Tournier, I., Marlin, R., Walton, K., Charbonnier, F., Coutant, S., Théry, J.-C., Charbonnier, C., Spurrell, C., Vezain, M., Ippolito, L., et al. (2014). Germline Mutations of Inhibins in Early-Onset Ovarian Epithelial Tumors. *Human Mutation* 35, 294–297.
- Tsang, W.P., Chau, S.P.Y., Kong, S.K., Fung, K.P., and Kwok, T.T. (2003). Reactive oxygen species mediate doxorubicin induced p53-independent apoptosis. *Life Sciences* 73, 2047–2058.
- Tutt, A., Tovey, H., Cheang, M.C.U., Kernaghan, S., Kilburn, L., Gazinska, P., Owen, J., Abraham, J., Barrett, S., Barrett-Lee, P., et al. (2018). Carboplatin in BRCA1/2 -mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. *Nature Medicine* 24, 628.

- Valle, L. (2014). Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World J Gastroenterol* 20, 9828–9849.
- Vassileva, V., Millar, A., Briollais, L., Chapman, W., and Bapat, B. (2002). Genes Involved in DNA Repair Are Mutational Targets in Endometrial Cancers with Microsatellite Instability. *Cancer Res* 62, 4095–4099.
- Vissers, L.E.L.M., Gilissen, C., and Veltman, J.A. (2016). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics* 17, 9.
- Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-Generation Sequencing: From Basic Research to Diagnostics. *Clinical Chemistry* 55, 641–658.
- Voer, R.M. de, Hahn, M.-M., Mensenkamp, A.R., Hoischen, A., Gilissen, C., Henkes, A., Spruijt, L., Zelst-Stams, W.A. van, Kets, C.M., Verwiel, E.T., et al. (2015). Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. *Scientific Reports* 5, 14060.
- Vogelstein, B., and Kinzler, K.W. (2004). Cancer genes and the pathways they control. *Nature Medicine* 10, 789–799.
- Vogelstein, B., Lane, D., and Levine, A.J. (2000). Surfing the p53 network.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. *Science* 339, 1546–1558.
- Walsh, C.S. (2015). Two decades beyond BRCA1/2: Homologous recombination, hereditary cancer risk and a target for ovarian cancer therapy. *Gynecologic Oncology* 137, 343–350.
- Walter, D.M., Venancio, O.S., Buza, E.L., Tobias, J.W., Deshpande, C., Gudiel, A.A., Kim-Kiselak, C., Cicchini, M., Yates, T.J., and Feldser, D.M. (2017). Systematic In Vivo Inactivation of Chromatin-Regulating Enzymes Identifies Setd2 as a Potent Tumor Suppressor in Lung Adenocarcinoma. *Cancer Res* 77, 1719–1729.
- Wang, D., Qian, X., Rajaram, M., Durkin, M.E., Lowy, D.R., Wang, D., Qian, X., Rajaram, M., Durkin, M.E., and Lowy, D.R. (2016). DLC1 is the principal biologically-relevant down-regulated DLC family member in several cancers. *Oncotarget* 7, 45144–45157.
- Wang, S., Konorev, E.A., Kotamraju, S., Joseph, J., Kalivendi, S., and Kalyanaraman, B. (2004). Doxorubicin Induces Apoptosis in Normal and Tumor Cells via Distinctly Different Mechanisms INTERMEDIACY OF H2O2- AND p53-DEPENDENT PATHWAYS. *J. Biol. Chem.* 279, 25535–25543.
- Wei, X., Walia, V., Lin, J.C., Teer, J.K., Prickett, T.D., Gartner, J., Davis, S., Program, N.C.S., Stemke-Hale, K., Davies, M.A., et al. (2011). Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genetics* 43, 442–446.
- Wells, S.A., Gosnell, J.E., Gagel, R.F., Moley, J., Pfister, D., Sosa, J.A., Skinner, M., Krebs, A., Vasselli, J., and Schlumberger, M. (2010). Vandetanib for the treatment of patients with locally advanced or metastatic hereditary medullary thyroid cancer. *J. Clin. Oncol.* 28, 767–772.
- Wilson, T.L.-S., Hattangady, N., Lerario, A.M., Williams, C., Koeppe, E., Quinonez, S., Osborne, J., Cha, K.B., and Else, T. (2017). A new POT1 germline mutation expanding the spectrum of POT1-associated cancers. *Familial Cancer* 16, 561–566.

- Wodarz, A., and Näthke, I. (2007). Cell polarity in development and cancer. *Nature Cell Biology* 9, 1016.
- Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. (2007). The Genomic Landscapes of Human Breast and Colorectal Cancers. *Science* 318, 1108–1113.
- Wyman, C., and Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet.* 40, 363–383.
- Xie et al. Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. - PubMed - NCBI.
- Xu, X., Hou, Y., Yin, X., Bao, L., Tang, A., Song, L., Li, F., Tsang, S., Wu, K., Wu, H., et al. (2012). Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor. *Cell* 148, 886–895.
- Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A., et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467, 1114–1117.
- Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *New England Journal of Medicine* 369, 1502–1511.
- Yeang, C.-H., McCormick, F., and Levine, A. (2008). Combinatorial patterns of somatic gene mutations in cancer. *The FASEB Journal* 22, 2605–2622.
- Yurgelun, M.B., Kulke, M.H., Fuchs, C.S., Allen, B.A., Uno, H., Hornick, J.L., Ukaegbu, C.I., Brais, L.K., McNamara, P.G., Mayer, R.J., et al. (2017). Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *JCO* 35, 1086–1095.
- Zhang, H., Head, P.E., Daddacha, W., Park, S.-H., Li, X., Pan, Y., Madden, M.Z., Duong, D.M., Xie, M., Yu, B., et al. (2016). ATRIP Deacetylation by SIRT2 Drives ATR Checkpoint Activation by Promoting Binding to RPA-ssDNA. *Cell Reports* 14, 1435–1447.
- Zhang, J., Liu, X., Scherer, D.C., Kaer, L. van, Wang, X., and Xu, M. (1998). Resistance to DNA fragmentation and chromatin condensation in mice lacking the DNA fragmentation factor 45. *PNAS* 95, 12480–12485.
- Zhang, Q., Davis, J.C., Lamborn, I.T., Freeman, A.F., Jing, H., Favreau, A.J., Matthews, H.F., Davis, J., Turner, M.L., Uzel, G., et al. (2009). Combined Immunodeficiency Associated with DOCK8 Mutations.
- Zhou, B.-B.S., and Elledge, S.J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature* 408, 433–439.
- Zhu, X., He, F., Zeng, H., Ling, S., Chen, A., Wang, Y., Yan, X., Wei, W., Pang, Y., Cheng, H., et al. (2014). Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nature Genetics* 46, 287.

Annexe 1: Gènes impliqués dans la prédisposition au cancer (Garber and Offit, 2005; Rahman, 2014; Vogelstein and Kinzler, 2004; Vogelstein et al., 2013)

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
ABCB11	Cholestase intrahépatique progressive familiale	Carcinome hépatocellulaire Cholangiocarcinome	Suppresseur de tumeur
APC	Polypose adénomateuse familiale	Cancer colorectal Hépatoblastome Tumeur desmoïde	Suppresseur de tumeur
ATM	Ataxie Téléangiectasie	<u>Mutations bi-alléliques:</u> Lymphomes, Leucémie <u>Mutation mono-allélique:</u> Cancer du sein	Suppresseur de tumeur
AXIN2	Syndrome de prédisposition au cancer-oligodontie	Cancer colorectal	Suppresseur de tumeur
BLM	Syndrome de Bloom	Lymphomes, Leucémie lymphoblastique aiguë Néoplasies myéloïdes hématologiques Carcinome épidermoïde Cancer gastrique, Cancer colorectal	Suppresseur de tumeur
BMPR1A	Polypose Juvénile	Cancer colorectal, Cancer gastrique, Hamartome	Suppresseur de tumeur
BRCA1	Syndrome sein-ovaire Anémie de Fanconi (mutations bialléliques)	Cancer du sein Cancer de l'ovaire	Suppresseur de tumeur
BRCA2	Syndrome sein-ovaire Anémie de Fanconi (mutations bialléliques)	<u>Mutations bi-alléliques:</u> Néoplasies myéloïdes hématologiques Medulloblastome Tumeur de Wilms <u>Mutations mono-alléliques:</u> Cancer du sein Cancer de l'ovaire Cancer de la prostate Cancer du pancréas	Suppresseur de tumeur
BRIP1	Anémie de Fanconi (mutations bialléliques)	<u>Mutations bi-alléliques:</u> Néoplasies myéloïdes hématologiques Carcinome épidermoïde <u>Mutations mono-alléliques:</u> Cancer du sein Cancer de l'ovaire	Suppresseur de tumeur
BUB1B	Syndrome d'aneuploïdie en mosaïque	Tumeur de Wilms Rhabdomyosarcome Néoplasies myéloïdes hématologiques	Suppresseur de tumeur
CBL	Syndrome de Noonan	Leucémie myélomonocytaire juvénile	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
CDC73	Syndrome d'hyperparathyroïdie-tumeur mandibulaire	Cancer de la parathyroïde Fibrome ossifiant	Suppresseur de tumeur
CDH1	Cancer gastrique diffus héréditaire	Cancer du sein (lobulaire) Cancer gastrique diffus	Suppresseur de tumeur
COL7A1	Epidermolyse bulleuse	Carcinome épidermoïde (peau)	Suppresseur de tumeur
CXCR4	Syndrome WHIM	Dysplasie génitale Cancer des muqueuses invasif	Oncogène
CYLD	Syndrome de Brooke-Spiegler	Cylindre Spiroadenocarcinome Carcinome basocellulaire	Suppresseur de tumeur
DDB2	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Melanome	Suppresseur de tumeur
DICER1	Syndrome DICER1	Blastome Pleuropulmonaire Néphrome kystique Tumeur maligne de Sertoli-Leydig de l'ovaire	Suppresseur de tumeur
DIS3L2	Syndrome de Perlman	Tumeur de Wilms	Suppresseur de tumeur
DKC1	Dyskératose congénitale	Leucémie myeloïde aiguë Carcinome épidermoïde	Suppresseur de tumeur
DOCK8	Syndrome hyper-IgE	Carcinome épidermoïde Lymphomes	Suppresseur de tumeur
ELANE	Neutropénie congénitale sévère	Leucémies	Suppresseur de tumeur
ERCC2	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Melanome	Suppresseur de tumeur
ERCC3	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Melanome	Suppresseur de tumeur
ERCC4	Xeroderma Pigmentosum Anémie de Fanconi	Carcinome basocellulaire Carcinome épidermoïde Melanome	Suppresseur de tumeur
ERCC5	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Melanome	Suppresseur de tumeur
FAH	Tyrosinémie	Carcinome hépatocellulaire	Suppresseur de tumeur
FANCA	Anémie de Fanconi	Néoplasies myeloïdes hématologiques Carcinomes épidermoïdes	Suppresseur de tumeur
FANCC	Anémie de Fanconi	Néoplasies myeloïdes hématologiques Carcinomes épidermoïdes	Suppresseur de tumeur
FANCD2	Anémie de Fanconi	Leucémie	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
FANCE	Anémie de Fanconi	Leucémie	Suppresseur de tumeur
FANCF	Anémie de Fanconi	Leucémie	Suppresseur de tumeur
FANCG	Anémie de Fanconi	Néoplasies myeloïdes hématologiques Carcinomes épidermoïdes	Suppresseur de tumeur
FAS	Syndrome lymphoprolifératif auto-immun	Lymphomes	Suppresseur de tumeur
FH	Léiomyomatose familiale et cancer du rein	Cancer du rein LeiomyoSarcome (uterus)	Suppresseur de tumeur
FLCN	Syndrome de Birt-Hogg-Dube	Cancer du rein Oncocytome	Suppresseur de tumeur
GATA2	Syndrome MonoMAC Syndrome d'Emberger	Néoplasies myeloïdes hématologiques	Suppresseur de tumeur
GBA	Maladie de Gaucher type 1	Myelome Lymphomes Carcinome hépatocellulaire	Suppresseur de tumeur
GJB2	Syndrome KID (Keratitits-Ichthyosis-Deafness)	Carcinome épidermoïde	Suppresseur de tumeur
GPC3	Syndrome de Simpson-Golabi-Behmel	Tumeur de Wilms Hepatoblastome, Carcinome hépatocellulaire Neuroblastome Gonadoblastome	Suppresseur de tumeur
HFE	Haemochromatose	Carcinome hépatocellulaire Cholangiocarcinome	Suppresseur de tumeur
HMBS	Porphyrie aiguë intermittente	Carcinome hépatocellulaire	Suppresseur de tumeur
HRAS	Syndrome de Costello	Rhabdomyosarcome Neuroblastome Carcinome urothélial (vessie)	Oncogène
ITK	Syndrome lymphoprolifératif	Lymphome de Hodgkin	Suppresseur de tumeur
LZTR1	Syndrome de Noonan Neurofibromatose type 3	Glioblastome Schwannome	Suppresseur de tumeur
MEN1	Néoplasie endocrinienne multiple type 1	Adénome hypophysaire et parathyroïdien Tumeur Neuroendocrine Tumeurs carcinoïdes Carcinome corticosurrénalien	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
MLH1	Syndrome de Lynch (mutation monoallélique) Syndrome de déficit constitutionnel de la réparation des mésappariements (mutation biallélique)	<u>Mutations bi-alléliques:</u> Tumeur du cerveau Hémopathies malignes Tumeurs embryonnaires <u>Mutations mono-alléliques:</u> Cancer colorectal Cancer de l'endomètre Cancer de l'ovaire	Suppresseur de tumeur
MPL	Thrombocytose familiale	Polypose adénomateuse familiale	Oncogène
MSH2	Syndrome de Lynch (mutation monoallélique) Syndrome de déficit constitutionnel de la réparation des mésappariements (mutation biallélique)	<u>Mutations bi-alléliques:</u> Tumeur du cerveau Hémopathies malignes Tumeurs embryonnaires <u>Mutations mono-alléliques:</u> Cancer colorectal Cancer de l'endomètre Cancer de l'ovaire Tumeurs sébacées	Suppresseur de tumeur
MSH6	Syndrome de Lynch (mutation monoallélique) Syndrome de déficit constitutionnel de la réparation des mésappariements (mutation biallélique)	<u>Mutations bi-alléliques:</u> Tumeur du cerveau Hémopathies malignes Tumeurs embryonnaires <u>Mutations mono-alléliques:</u> Cancer colorectal Cancer de l'endomètre Cancer de l'ovaire	Suppresseur de tumeur
MTAP	Syndrome de sténose médullaire diaphysaire-tumeur osseuse	Histiocytome fibreux	Suppresseur de tumeur
NBN	Syndrome de Nijmegen	Lymphomes Medulloblastome Gliome Rhabdomyosarcome	Suppresseur de tumeur
NF1	Neurofibromatose type 1	Gliome Tumeur maligne des gaines nerveuses périphériques	Suppresseur de tumeur
NF2	Neurofibromatose type 2	Schwannome vestibulaire Meningiome Ependymome	Suppresseur de tumeur
NTHL1	Polypose adénomateuse familiale	Cancer colorectal	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
PALB2	Anémie de Fanconi	<u>Mutations bi-alléliques:</u> Néoplasies myeloïdes hématologiques Medulloblastome Neuroblastome Tumeur de Wilms <u>Mutations mono-alléliques:</u> Cancer du sein Cancer du pancreas	Suppresseur de tumeur
PMS1	Syndrome de Lynch	Cancer colorectal Cancer de l'ovaire Cancer de l'endomètre	Suppresseur de tumeur
PMS2	Syndrome de Lynch (mutation monoallélique) Syndrome de déficit constitutionnel de la réparation des mésappariements (mutation biallélique)	<u>Mutations bi-alléliques:</u> Tumeur du cerveau Hémopathies malignes Tumeurs neuroectodermiques primitives <u>Mutations mono-alléliques:</u> Cancer colorectal Cancer de l'endomètre Cancer de l'ovaire	Suppresseur de tumeur
POLD1	Polypose adénomateuse associée aux ADN polymérases	Cancer colorectal Cancer de l'endomètre	Suppresseur de tumeur
POLE	Polypose adénomateuse associée aux ADN polymérases	Cancer colorectal	Suppresseur de tumeur
POLH	Xeroderma Pigmentosum	Carcinome épidermoïde	Suppresseur de tumeur
PRKAR1A	Complexe de Carney	Myxome (cardiaque/cutané/sein) Cancer de la thyroïde Tumeur stromale du cordon sexuel	Suppresseur de tumeur
PTCH1	Syndrome de Gorlin	Carcinome basocellulaire Medulloblastome	Suppresseur de tumeur
PTEN	Syndrome de Cowden	Cancer du sein Cancer de la thyroïde Cancer de l'endomètre	Suppresseur de tumeur
PTPN11	Syndrome de Noonan	Leucémie myelomonocytaire juvénile Neuroblastome	Oncogène
PTPN13	Carcinome hépatocellulaire héréditaire	Carcinome hépatocellulaire	Suppresseur de tumeur
RAD51C	Anémie de Fanconi	Mutation mono-allélique: Cancer de l'ovaire	Suppresseur de tumeur
RECQL4	Syndrome de Rothmund-Thompson	Osteosarcome Carcinome basocellulaire Carcinome épidermoïde	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
RET	Néoplasie endocrinienne multiple type 2A/2B Carcinome médullaire thyroïdien	Cancer médullaire de la thyroïde Pheochromocytome	Oncogène
RFWD3	Anémie de Fanconi	Leucémie myéloïde aiguë Tumeurs de la tête et du cou	Suppresseur de tumeur
RMRP	Hypoplasie cartilage-cheveux	Lymphomes non Hodgkinniens Carcinome épidermoïde Leucémies	Suppresseur de tumeur
SBDS	Syndrome de Schwachman-Diamond	Néoplasies myéloïdes hématologiques	Suppresseur de tumeur
SDHA	Syndrome de Carney-Stratakis	Parangliome Pheochromocytome Tumeur stromale gastrointestinale	Suppresseur de tumeur
SDHAF2	Phéochromocytome-parangliome héréditaire	Parangliome Pheochromocytome	Suppresseur de tumeur
SDHB	Phéochromocytome-parangliome héréditaire	Parangliome Pheochromocytome Cancer du rein	Suppresseur de tumeur
SDHC	Phéochromocytome-parangliome héréditaire	Parangliome Pheochromocytome Tumeur stromale gastrointestinale	Suppresseur de tumeur
SDHD	Phéochromocytome-parangliome héréditaire	Parangliome Pheochromocytome Tumeur stromale gastrointestinale	Suppresseur de tumeur
SERPINA1	Déficit en alpha-1-antitrypsine	Carcinome hépatocellulaire	Suppresseur de tumeur
SETBP1	Syndrome de Schinzel-Giedion	Tumeurs neuroépithéliales	Oncogène
SH2D1A	Maladie lymphoproliférative liée à l'X	Lymphomes	Suppresseur de tumeur
SLC25A13	Citrullinémie de type II	Carcinome hépatocellulaire	Suppresseur de tumeur
SMAD4	Polypose Juvénile	Cancer colorectal	Suppresseur de tumeur
SMARCA4	Tumeur rhabdoïde familiale	Tumeur Rhabdoïde	Suppresseur de tumeur
SMARCB1	Tumeur rhabdoïde familiale	Tumeur Rhabdoïde Tumeur neuroectodermique primitive	Suppresseur de tumeur
SOS1	Syndrome de Noonan	Rhabdomyosarcome	Oncogène
STAT3	Syndrome hyper-IgE	Lymphomes	Oncogène
STK11	Syndrome de Peutz-Jeghers	Cancer colorectal Cancer gastrique Cancer du sein Tumeur stromale du cordon sexuel	Suppresseur de tumeur
TERT	Dyskératose congénitale	Leucémie myéloïde aiguë Carcinome épidermoïde Mélanome	Suppresseur de tumeur

Gène	Syndrome avec prédisposition au cancer	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer syndromique			
TGFBR1	Syndrome de Ferguson-Smith	Carcinome épidermoïde (peau)	Suppresseur de tumeur
TGFBR2	Syndrome de Lynch	Cancer colorectal	Suppresseur de tumeur
TP53	Syndrome de Li-Fraumeni	Cancer du sein Sarcome Carcinome corticosurrénalien Astrocytome	Suppresseur de tumeur
TRIM37	Nanisme Mulibrey	Tumeur de Wilms	Suppresseur de tumeur
TSC1	Sclérose tubéreuse de Bourneville	Cancer du rein, angiomyolipome Astrocytome sous épendymaire à cellules géantes Rhabdomyome (cardiac)	Suppresseur de tumeur
TSC2	Sclérose tubéreuse de Bourneville	Cancer du rein, angiomyolipome Astrocytome sous épendymaire à cellules géantes Rhabdomyome (cardiaque)	Suppresseur de tumeur
UROD	Porphyrie cutanée tardive familiale	Carcinome hépatocellulaire	Suppresseur de tumeur
VHL	Syndrome de Von Hippel-Lindau	Cancer du rein Pheochromocytome Tumeur Neuroendocrine (pancreas) Hémangioblastome (système nerveux central, rétine)	Suppresseur de tumeur
WAS	Syndrome de Wiskott-Aldrich	Lymphomes	Oncogène
WRN	Syndrome de Werner	Sarcome Mélanome Cancer de la thyroïde	Suppresseur de tumeur
WT1	Syndrome WAGR (<i>Wilms tumor--Aniridia--Genitourinary anomalies--mental Retardation</i>) Syndrome de Denys-Drash Syndrome de Frasier	Tumeur de Wilms Gonadoblastome	Suppresseur de tumeur
XPA	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Mélanome	Suppresseur de tumeur
XPC	Xeroderma Pigmentosum	Carcinome basocellulaire Carcinome épidermoïde Mélanome	Suppresseur de tumeur

Gène	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer non syndromique		
ALK	Neuroblastome	Oncogène
APOBEC3B	Cancer du sein	Suppresseur de tumeur
AR	Cancer de la prostate	Oncogène
ATR	Cancer oropharyngé	Suppresseur de tumeur
BAP1	Melanome (cutaneous, uveal) Mesotheliome Meningiome Cancer du poumon	Suppresseur de tumeur
BARD1	Cancer de l'ovaire Cancer du sein Cancer de l'endomètre	Suppresseur de tumeur
CDH10	Adénocarcinome canalaire du pancréas	Suppresseur de tumeur
CDK4	Melanome	Oncogène
CDKN1B	Cancer de la thyroïde, Adénome hypophysaire	Suppresseur de tumeur
CDKN2A	Melanome [p16 and p14ARF] Cancer du pancreas [p16] Astrocytome [p14ARF]	Suppresseur de tumeur
CEBPA	Néoplasies myeloïdes hématologiques	Suppresseur de tumeur
CHEK2	Cancer du sein	Suppresseur de tumeur
EGFR	Cancer du poumon non à petites cellules	Oncogène
EXT1	Chondrosarcome	Suppresseur de tumeur
EXT2	Chondrosarcome	Suppresseur de tumeur
FAT1	Cancer du pancreas	Suppresseur de tumeur
FEN1	Cancer du sein	Suppresseur de tumeur
KDR	Melanome	Oncogène
KIT	Tumeur stromale gastro-intestinale	Oncogène
LMO1	Neuroblastome	Oncogène
MAX	Paragangliome Pheochromocytome	Suppresseur de tumeur
MET	Cancer du rein (carcinome papillaire)	Oncogène
MUTYH	Cancer colorectal	Suppresseur de tumeur
PDGFRA	Tumeur stromale gastro-intestinale	Oncogène
PHOX2B	Neuroblastome	Suppresseur de tumeur
POLQ	Cancer du sein	Suppresseur de tumeur
PRF1	Leucémies Lymphomes	Suppresseur de tumeur
PRSS1	Cancer du pancreas	Suppresseur de tumeur
RAD51D	Cancer de l'ovaire	Suppresseur de tumeur
RB1	Retinoblastome Pinealome Sarcome Melanome	Suppresseur de tumeur

Gène	Principale type de tumeur associé	Classe
Gènes associés à une prédisposition au cancer non syndromique		
RHBDF2	Cancer de l'oesophage	Oncogène
RUNX1	Néoplasies myéloïdes hématologiques(Leucémies)	Suppresseur de tumeur
SMARCE1	Meningiome	Suppresseur de tumeur
SPOP	Cancer de la prostate	Suppresseur de tumeur
SRY	Gonadoblastome	Suppresseur de tumeur
SUFU	Medulloblastome, Meningiome	Suppresseur de tumeur
TMEM127	Pheochromocytome	Suppresseur de tumeur
TP63	Cancer du poumon non à petites cellules	Suppresseur de tumeur
TSHR	Cancer de la thyroïde	Oncogène

Annexe 2: Liste des amorces

Amorce	Séquence	Température d'hybridation	Taille amplicon	% GC	Coordonnées amplicon (hg19)
ATR					
Séquençage haut débit ciblé					
ATR_ex1_F	CGGGTCTATGCAGAAAAGAC	60.99	410	66	chr3:142297311-142297720
ATR_ex1_R	TATCAACGGGCACTCTGAGG	61.20			
ATR_ex2-5_F	GAGAAGTTCAGGTCTAGGCTTTG	58.73	7276	36	chr3:142279958-142287233
ATR_ex2-5_R	CCCTTGCTACATTTAGAAAGATG	60.36			
ATR_ex6-12_F	GTGGCTAAGTTGCCATATTTTCAG	60.04	7660	37	chr3:142272114-142279773
ATR_ex6-12_R	AACACTTTTAGCTGCAACCAGAG	60.00			
ATR_ex13-16_F	GAAGGTATACTTGGCAGCATTTG	60.04	6196	37	chr3:142266290-142272485
ATR_ex13-16_R_2	TCAATGCTAATATATGTTGTTAATGCC	59.64			
ATR_ex17-20_F	CATGTGAAGATATATGCTTTTGGAG	59.12	6985	33	chr3:142254782-142261766
ATR_ex17-20_R	AAAAGTCATGGATTTCATGGTAGC	59.39			
ATR_ex21_F	GCTGTGGGATGATTCTAGTTTTG	60.01	1313	32	chr3:142253305-142254617
ATR_ex21_R	AAATTCATGAGGGTCTTAGGCTC	59.98			
ATR_ex22-24_F	AATCTTCACTGGCTTGAGAATTG	59.78	5710	36	chr3:142238014-142243723
ATR_ex22-24_R	AGAATTTAACCCAATGAGCAGAC	58.71			
ATR_ex25-27_F	AGCTAGACACTGAAGTCACCCAG	59.99	4153	37	chr3:142230625-142234777
ATR_ex25-27_R	TTCTCAGTGAAAAGGCCACAC	60.28			
ATR_ex28-30_F	AGTTGGAGCAAGACTCTCTGAAC	59.20	5115	37	chr3:142222013-142227127
ATR_ex28-30_R	GACAGAAGCTGTTAAGTGGTAGACTG	59.94			
ATR_ex31-34_F	GAGTTATTGGAATTGCTTACCTCTG	59.61	4005	36	chr3:142214958-142218962
ATR_ex31-34_R	CCTTGAAATGGGAAGGTATCTG	59.83			
ATR_ex35_F	AACCTGTGGCTACAGAGAGCC	60.83	639	31	chr3:142211743-142212381
ATR_ex35_R	ATTACTGGGATGAAGGGTAGTGG	60.47			
ATR_ex36_F	GCTATTCTTTTCCACATCAAAG	59.18	571	36	chr3:142203711-142204281
ATR_ex36_R	ATGATTGCTGTGTTGGTGTGAG	61.02			
ATR_ex36_F2	TTTCCAAAACACATGTCTATATCAC	57.24	498	36	chr3:142203761-142204258
ATR_ex36_R2	ATTGAGTCTGCTTTTGACTTTGC	59.95			
ATR_ex37-39_F	TTTCCCAGAAATAGAATGATTGG	59.34	2871	31	chr3:142186525-142189395
ATR_ex37-39_R	TCTCGGAAAATTCTTAATTCACAC	58.74			
ATR_ex40-41_F	ATGACAGAGGTTAAAATGGGATG	59.25	2032	32	chr3:142183493-142185524
ATR_ex40-41_R	GGTACCTCAAATGGTTAAATTAATGG	60.12			
ATR_ex42-45_F	AAAAGAAGGACCTACCCAGGAC	59.87	5035	33	chr3:142176126-142181160
ATR_ex42-45_R	TAGATAGCAGGGGTGGGAGG	61.35			
ATR_ex46-47_F	ATGGCACATAGAACCCATCTTAC	59.28	4335	40	chr3:142167997-142172331
ATR_ex46-47_R	GACTTGCTTGTCTTGCAAATATAG	59.78			
Séquençage Sanger					
ATR_ex32_F	TCAGGTGACATTTATAGGCCAGA	58.96	339	37	chr3:142217389-142217727
ATR_ex32_R	GGAATGAAGCCACCTGTTTGATA	58.98			
ATR_ex47F	AATTAGACTTGCTTGTCTTCTGC	57,02	682	31	chr3:142167992-142168673
ATR_ex47R	TCCAGAAACATGTCCAAGCAG	61,22			
Séquençage haut débit ciblé sur cDNA					
ATR_cDNA_ex46-47F	TACATGATCCTCTTGTTGGAATGG	61.10	393	37	chr3:142168141-142172058
ATR_cDNA_ex46-47R	TCAGAGAGAAATAACAGTTGCTGAG	59.27			

Amorce	Séquence	Temperature d'hybridation	Taille amplicon	% GC	Coordonnées amplicon (hg19)
DLC1					
Séquençage haut débit ciblé					
DLC1_ex1altF	TCACTCCTCCAAAATTCAAACTC	59.62	1022	68	chr8:12990241-12991262
DLC1_ex1altR	TACGCAGAGCTACCAAGAAAAG	60.08			
DLC1_ex2F	GTATCACTTAAAATGGCACCAGC	59,92	2053	37	chr8:13356023-13358075
DLC1_ex2R	GTAAATCAGGAAAACATGCAAGC	60.02			
DLC1_ex3F	GCTTGCTAACAGAGTCTCTCTCG	59.98	1011	38	chr8:13258646-13259656
DLC1_ex3R	TTAAGTAGCAGTGGCTAAATGGC	59.83			
DLC1_ex4F	TTGATTCATGACCTGTTTCAGTG	60,02	1206	39	chr8:13250845-13252050
DLC1_ex4R	TGCAACCAAACCACCTATAACTC	60,28			
DLC1_ex5F	TGTCATTATGTCAGAAAACCACG	59.91	1209	37	chr8:13162582-13163790
DLC1_ex5R	TTGGAAAGATCTCATTCTGAACC	59.58			
DLC1_ex6-7F	CTATGATCACTCTATGTTTGCC	57.79	5209	39	chr8:12968115-12973323
DLC1_ex6-7R	GTAACCATTCGAATTAAGCGTG	59.92			
DLC1_ex8-10F	ACTCCACTGGCTTCAACAATATC	59.54	5578	47	chr8:12955362-12960939
DLC1_ex8-10R	GTATCATAGCTGTGTGCCATTTT	58.67			
DLC1_ex11-15F	CAGATCCCCAGACAGATTAGTTG	60.00	5506	45	chr8:12947661-12953166
DLC1_ex11-15R	TTCTATACCGCCATTCTGTAAG	59.55			
DLC1_ex16-18F	AATACAGACCTCAACAAACAGG	59.44	3279	43	chr8:12943135-12946413
DLC1_ex16-18R	GGGACAGCAGCCTATTTAAGTC	59.29			
Séquençage Sanger					
DLC1_ex9F	GAAGCCCAACCCAGAATC	60.68	473	58	chr8:12956874-12957346
DLC1_ex9R	GGCTCTGTGAACTGGAGGAC	60.04			
DLC1_ex10F	GAGTGATTGCCATAGGATGACAG	60.89	692	45	chr8:12955771-12956462
DLC1_ex10R	CCATTGCTAATGTTGCTTCTACC	60.04			
Séquençage haut débit ciblé sur cDNA					
DLC1_cDNA_ex9-12F	AAGCAGACTGTCTTCGTAGTTG	59.99	874	56	chr8:12952364-12957289
DLC1_cDNA_ex9-12R	AGGGAAAACAGTAGCGACAGC	60.81			
DLC1_cDNA_ex9-11F	AGTCTGGAACCTTGATCCTCTTC	60.12	366	53	chr8:12952714-12957003
DLC1_cDNA_ex9-11R	AACAGATACACCTGGATGTGGAC	60.17			
TP53I3					
Séquençage haut débit ciblé					
TP53I3_F	CGGAATATGTGGGGTTCTGG	62.40	8099	50	chr2:24300119-24308217
TP53I3_R	ACGGATACAGCTTTTACCCAATC	60.59			
Séquençage Sanger					
TP53I3_ex1F	GTCCACTGAGTGCTGTTGAGAG	60.10	552	68	chr2:24306989-24307540
TP53I3_ex1R	GAGTGTGATTGCTGATCCTC	61.03			
TP53I3_ex1bF	GTACCTGCATTAAGTCCGCC	61.08	150	61	chr2:24307055-24307204
TP53I3_ex1bR	CAGACAATATGTTAGCCGTGCAC	60.24			
TP53I3_ex1cF	CACCTTCAGGAGGACTTCACC	61.07	104	60	chr2:24307095-24307198
TP53I3_ex1cR	ATATGTTAGCCGTGCACTTTGAC	60.44			
BLM					
Séquençage Sanger					
BLM_ex8F	GTAAGTAGCAACTGGGCTGAAAC	59.38	209	36	chr15:91304104-91304312
BLM_ex8R	ATGTCTCCAGTCATCATCATC	60.21			

Amorce	Séquence	Temperature d'hybridation	Taille amplicon	% GC	Coordonnées amplicon (hg19)
BLM_ex10F	ATTTGAGTATGGCAAATTGTTGG	60.12	381	36	chr15:91308389-91308769
BLM_ex10R	ACTGCAAATTTAACTGCTGTGC	59.47			
BIK					
Séquençage Sanger					
BIK_ex4F	GAATGGGGGTTTATTTCCAC	59.41	556	58	chr22:43524206-43524761
BIK_ex4R	AGGCAGATGTGGACACAGG	59.76			
Séquençage haut débit ciblé sur cDNA					
BIK_cDNA_ex2-5F	CCTATGGAGGACTTCGATTCTTT	59.98	470	61	chr22:43520149+43525418
BIK_cDNA_ex2-5R	AAAAAGGCATCATGAGAAAACAG	59.66			
DFFA					
Séquençage Sanger					
DFFA_ex5F	AACTGGCAGCTAGGATTCACAG	60.81	438	56	chr1:10522992-10523429
DFFA_ex5R	TGGGAGGACACTATTGTGATG	59.85			
SETD2					
Séquençage Sanger					
SETD2_ex8F	CCTTAAGGGTCAGAAGTGTCATAC	58.58	785	33	chr3:47142769-47143553
SETD2_ex8R	CTATACTCAGCCAGATGACCACC	59.99			

ÉCOLE PRATIQUE DES HAUTES ÉTUDES SCIENCES DE LA VIE ET DE LA TERRE

Recherche des bases moléculaires des phénotypes extrêmes de cancer par séquençage d'exome

Martin Chevarin

Certains cancers peuvent être qualifiés de « phénotypes extrêmes ». Il s'agit soit de formes sporadiques particulièrement précoces, soit de formes familiales avec un excès de cancers dans une même branche parentale. La présence de plusieurs tumeurs primitives, de tumeurs bilatérales ou l'association de plusieurs cancers rares chez un même patient ou au sein d'une même famille, peut également entrer dans cette catégorie. Devant ces présentations, une prédisposition génétique est très fortement suspectée.

L'analyse de panels de gènes connus pour être impliqués dans les formes mendéliennes de cancer est la pratique courante pour identifier la mutation responsable d'une prédisposition au cancer. Cependant, elle ne permet d'identifier la mutation responsable de la prédisposition génétique que dans 10% des cas. Il reste donc de nombreuses cas avec une prédisposition génétique au cancer indiscutable mais sans diagnostic moléculaire. Cela représente une difficulté majeure pour les oncogénéticiens souhaitant proposer une surveillance et un conseil génétique des plus adaptés aux patients et à leurs apparentés.

Le séquençage de l'exome, de 27 patients présentant des phénotypes extrêmes de cancer, a permis d'identifier 9 gènes candidats pour une prédisposition génétique au cancer :

- Trois de ces gènes candidats (*ATR*, *DLC1* et *TP53/3*) ont été retrouvés à partir de cas familiaux de cancers du sein masculin, un cancer particulièrement rare. Le séquençage ciblé de ces gènes dans une cohorte de réplication a permis d'identifier des récurrences pour *ATR* et *DLC1* et des analyses fonctionnelles ont permis de préciser la pathogénicité des certaines de ces variations.

- Deux variations ont été identifiées dans des gènes déjà connus comme responsables de prédisposition génétique au cancer. L'une d'elle a été identifiée dans le gène *PALB2* avec une localisation de cancer pulmonaire inhabituelle, diagnostiqué à 26 ans. La seconde est une variation hétérozygote du gène *BLM* dans un cancer de l'ovaire à 30 ans.

- Les quatre autres variations candidates concernent des gènes impliqués dans les processus d'apoptose (*BIK* et *DFFA*), de régulation épigénétique (*SETD2*) ou de régulation de la prolifération cellulaire (*FAT4*).

Ces résultats montrent la pertinence d'utiliser le séquençage d'exome et une approche de phénotypes extrêmes pour identifier des variations rares dans de nouveaux gènes candidats de prédisposition au cancer.

Mots-clé : Cancer, Prédisposition, Exome, Phénotype extrême, *ATR*